

us-09-864-291-11.rge

Wed Dec 17 07:08:14 2003

JOURNAL Patent: WO 0190185-A 11 29-NOV-2001;
QUEEN'S UNIVERSITY AT KINGSTON (CA); OREGON HEALTH SCIENCES
UNIVERSITY (US)

FEATURES
source Location/Qualifiers
1..1001
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

CDS
1..708
/note="unannotated protein product"

/codon_start=1
/protein_id="CAD33337.1"
/db_xref="GI:18675412"
translation="MPPDLMTNLTVEOPVPAANFKITQIAAPYGMGEGAFPLVER
/transl_except="MPPDLMTNLTVEOPVPAANFKITQIAAPYGMGEGAFPLVER
NDAIAPAPVPAANFKITQIAAPYGMGEGAFPLVER
APPAPVPAANFKITQIAAPYGMGEGAFPLVER
PLGAPVPAANFKITQIAAPYGMGEGAFPLVER
PENAPVPAANFKITQIAAPYGMGEGAFPLVER"

BASE COUNT 261 a 254 c 248 g 238 t

ORIGIN
Query Match 100.0%; Score 1001; DB 6; Length 1001;
Best Local Similarity 100.0%; Pred. No. 7.4e-259; Indels 0; Gaps 0;
Matches 1001; Conservative 0; Mismatches 0

1 ATGCCATTGATCTGATGACGAACTCCTGTTGAAACAGATTTTGTCTGCAACTTC 60
1 ATGCCATTGATCTGATGACGAACTCCTGTTGAAACAGATTTTGTCTGCAACTTC 60
61 ATTTAGGAACTATTCAAGCAGCTCATATGTTGGTGGAGAGACAGCTCTTTTAA 120
61 ATTTAGGAACTATTCAAGCAGCTCATATGTTGGTGGAGAGACAGCTCTTTTAA 120
61 ATTTAGGAACTATTCAAGCAGCTCATATGTTGGTGGAGAGACAGCTCTTTTAA 120
121 TTAGCTTCAGAAATGAGATGTCATGAAATTTGCCAGTTGATGTTGAAAGCTGCTCT 180
121 TTAGCTTCAGAAATGAGATGTCATGAAATTTGCCAGTTGATGTTGAAAGCTGCTCT 180
121 TTAGCTTCAGAAATGAGATGTCATGAAATTTGCCAGTTGATGTTGAAAGCTGCTCT 180
121 TTAGCTTCAGAAATGAGATGTCATGAAATTTGCCAGTTGATGTTGAAAGCTGCTCT 180
181 GCTGTGGCCGAGAGATTTCACTTGAACCTTAAATGAGCTTCACTTATGGAAT 240
181 GCTGTGGCCGAGAGATTTCACTTGAACCTTAAATGAGCTTCACTTATGGAAT 240
181 GCTGTGGCCGAGAGATTTCACTTGAACCTTAAATGAGCTTCACTTATGGAAT 240
241 TATGTAATTAATGAGGAGAAATATGTCATGCAAGATGCTTGTGAGTATGTC 300
241 TATGTAATTAATGAGGAGAAATATGTCATGCAAGATGCTTGTGAGTATGTC 300
241 TATGTAATTAATGAGGAGAAATATGTCATGCAAGATGCTTGTGAGTATGTC 300
241 TATGTAATTAATGAGGAGAAATATGTCATGCAAGATGCTTGTGAGTATGTC 300
301 TATGTAATTAATGAGGAGAAATATGTCATGCAAGATGCTTGTGAGTATGTC 300
301 TATGTAATTAATGAGGAGAAATATGTCATGCAAGATGCTTGTGAGTATGTC 300
301 TATGTAATTAATGAGGAGAAATATGTCATGCAAGATGCTTGTGAGTATGTC 300
301 TATGTAATTAATGAGGAGAAATATGTCATGCAAGATGCTTGTGAGTATGTC 300
361 GATATGAGAGCCCAACCGTGAATGAGAGCCCGCTGAGGATACAGAGCTCACT 420
361 GATATGAGAGCCCAACCGTGAATGAGAGCCCGCTGAGGATACAGAGCTCACT 420
361 GATATGAGAGCCCAACCGTGAATGAGAGCCCGCTGAGGATACAGAGCTCACT 420
361 GATATGAGAGCCCAACCGTGAATGAGAGCCCGCTGAGGATACAGAGCTCACT 420
421 GTGCGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
421 GTGCGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
421 GTGCGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
421 GTGCGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
481 CCTCTAGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
481 CCTCTAGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
481 CCTCTAGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
481 CCTCTAGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
541 CCACTCTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
541 CCACTCTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
541 CCACTCTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
541 CCACTCTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
601 GCTCACTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
601 GCTCACTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
601 GCTCACTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
601 GCTCACTGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCC 540
661 AACGAGGCTTCTTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
661 AACGAGGCTTCTTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
661 AACGAGGCTTCTTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
661 AACGAGGCTTCTTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720

721 TAACTCTGAGAGCTCAACGAGAGAGGTAACCTTAAATTTGAGTCAGATAGAGAG 780
721 TAACTCTGAGAGCTCAACGAGAGAGGTAACCTTAAATTTGAGTCAGATAGAGAG 780
721 TAACTCTGAGAGCTCAACGAGAGAGGTAACCTTAAATTTGAGTCAGATAGAGAG 780
781 AGGACTGAGCTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 840
781 AGGACTGAGCTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 840
781 AGGACTGAGCTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 840
841 CCAATGCTTCAAGAGAGAGCTCAACCACTAGAGAGGATAGGTTCCCTGTTCACTTTC 960
841 CCAATGCTTCAAGAGAGAGCTCAACCACTAGAGAGGATAGGTTCCCTGTTCACTTTC 960
841 CCAATGCTTCAAGAGAGAGCTCAACCACTAGAGAGGATAGGTTCCCTGTTCACTTTC 960
901 GAAAGAGAGAGAGCTCAACCACTAGAGAGGATAGGTTCCCTGTTCACTTTC 960
901 GAAAGAGAGAGAGCTCAACCACTAGAGAGGATAGGTTCCCTGTTCACTTTC 960
901 GAAAGAGAGAGAGCTCAACCACTAGAGAGGATAGGTTCCCTGTTCACTTTC 960
961 AGTGTCTGCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1001
961 AGTGTCTGCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1001
961 AGTGTCTGCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1001

RESULT 2
BC022546 2267 bp mRNA 11near PRI 04-FEB-2002
LOCUS Homo sapiens, similar to RIKEN CDNA 4930521123 gene, clone
DEFINITION MGC:26816 IMAGE:4811804, mRNA, complete cds.
ACCESSION BC022546 GI:18490706
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Tissue: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 2267)
Struhsberg, R.
Direct Submission
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcrp@nci.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcd@pdx1.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

REMARK
COMMENT
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/INL at: http://image.lnl.gov
Series: IRAX Plate: 32 Row: 3 Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
Location/Qualifiers
1..2267
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26816 IMAGE:4811804"
/tissue_type="Brain, hippocampus"
/clone_id="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBlueScript"
32..961
/codon_start=1
/product="Similar to RIKEN CDNA 4930521123 gene"

CDS


```

/protein_id="AAH22546.1"
/db_xref="gi:18490707"
/translation="MAVQSHTEENRRGALLINGESILKRSNPVELSPFORSGSNVRS
GRKGTLELTISYRVIPLTSCSISDPMFSFMPPLDNLTLVEQVPANPLKTTIOAA
PYGMEGQATPKLIFRNGGALISFQALWKASAAARGPLRLNDWSSKGIYVITBS
GNMCTPQMFCEVIVYGAAPAGYGAAPPGYGAAPAGYGAQPGNNGSPFVGYASVRRG
APPLGYGAAPAGYGAAPPLGYGAAPPLGYGAPPLGYGAPPLGYGAPPLGAGPPAGYR
PAGGAPDHSTAAQAPENASLPSASSQVHS"

```

	Query Match	Similarity	98.0%	Score	783.2	DB	9	Length	2267	
	Best Local	Similarity	99.0%	Pred.	No.	4.7e-200				
	Matches	788	Conservative	0	Mismatches	8	Indels	0	Gaps	0
QY	1	ATGCCATTGTAATCTGATGACGAACTCACTGTTGAACAACAGATATTTGCTGCAAACTTC	60							
Db	254	ATGCCATTGTAATCTGATGACGAACTCACTGTTGAACAACAGATATTTGCTGCAAACTTC	313							
QY	61	ATTAAAGGAACTAATCAGGAGCTCCATATGCTGCTGGAAGGCAAGCTACTTTTAA	120							
Db	314	ATTAAAGGAACTAATCAGGAGCTCCATATGCTGCTGGAAGGCAAGCTACTTTTAA	373							
QY	121	TTAGCTTCAGAAATGAGATGCAATTGAATTTGCCAGTTGATGTTAAAGTCCCTCT	180							
Db	374	TTAGCTTCAGAAATGAGATGCAATTGAATTTGCCAGTTGATGTTAAAGTCCCTCT	433							
QY	181	GCTGTGCTGCAAGGATTTTCCACTTATGAACCTTAAATGACTGCTTATGGAAT	240							
Db	434	GCTGTGCTGCAAGGATTTTCCACTTATGAACCTTAAATGACTGCTTATGGAAT	493							
QY	241	TATGTAAATTAATGGAAGGAAATATGCACTCCAGATGCTGTTGACGTTATGTC	300							
Db	494	TATGTAAATTAATGGAAGGAAATATGCACTCCAGATGCTGTTGACGTTATGTC	553							
QY	301	TATGAGGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCA	360							
Db	554	TATGAGGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCA	613							
QY	361	GGATATGAGGCCCACTGAGATATGAGCCCGCTGTGGAATACAGAGCTTCACT	420							
Db	614	GGATATGAGGCCCACTGAGATATGAGCCCGCTGTGGAATACAGAGCTTCACT	673							
QY	421	GTGCGATATGAGGCCCACTTGTGATACGAGGCCCACTGCAAGATATGAGCCCA	480							
Db	674	GTGCGATATGAGGCCCACTTGTGATACGAGGCCCACTGCAAGATATGAGCCCA	733							
QY	481	CCTCTAGATATGAGGCCCACTTGTGATATGAAACCCCACTCTGATATGAGCC	540							
Db	734	CCTCTAGATATGAGGCCCACTTGTGATATGAAACCCCACTCTGATATGAGCC	793							
QY	541	CCACTCTCGGATATGAGGCCCACTGCAAGAAATGAAGCCCGGCTGCGGATACGA	600							
Db	794	CCACTCTCGGATATGAGGCCCACTGCAAGAAATGAAGCCCGGCTGCGGATACGA	853							
QY	601	GCTCTACCTGCTGATACAGAGCCAGGCTCAGAGATCTACAGAGCCCAAGCTCTGA	660							
Db	854	GCTCTACCTGCTGATACAGAGCCAGGCTCAGAGATCTACAGAGCCCAAGCTCTGA	913							
QY	661	AACGAGCTTCTCTTCCCTCTGCTCTCTTCTCAAGTCAATTCTTAACTTTCAAGTG	720							
Db	914	AACGAGCTTCTCTTCCCTCTGCTCTCTTCTCAAGTCAATTCTTAACTTTCAAGTG	973							
QY	721	TAAACCTTGAAGCTCAACCAAGAGGTAACTTAAATTTGAAGTCAAGATTAAGAG	780							
Db	974	TAAACCTTGAAGCTCAACCAAGAGGTAACTTAAATTTGAAGTCAAGATTAAGAG	1033							
QY	781	ACGACTCAGCTTAAG	796							
Db	1034	ACGACTCAGCTATG	1049							

BC022549	BC022549	2266 bp	mRNA	linear	PRI 04-FEB-2002
LOCUS					
DEFINITION	Homo sapiens, similar to RIKEN CDNA 4930521i23 gene, clone				
	MGC:26828 IMAGE:4815849, mRNA, complete cdb.				
ACCESSION	BC022549				
VERSION	BC022549.1	GI:18490710			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
REFERENCE	1 (bases 1 to 2266)				
AUTHORS	Strausberg, R.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-FEB-2002) National Institutes of Health, Mammalian				
	Gene Collection (MGC), Cancer Genomics Office, National Cancer				
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,				
	USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk				

Tissue Procurement: Mikiros Palkovits, M.D., Ph.D.
 DNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-hsbc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxill.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.gov>
 Series: IRAX Plate: 32 Row: m Column: 3
 This clone was selected for full length sequencing because it
 passed the following selection criteria: similarity but not
 identity to protein.

FEATURES	SOURCE	Location/Qualifiers
		1..2266
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="MGC:26828 IMAGE:4815849"
		/tissue_type="Brain, hippocampus"
		/clone_id="NH_MGC_95"
		/lab_host="DH10B"
		/note="Vector: pBluescript"
CDS		32..634
		/codon_start=1
		/product="Similar to RIKEN cDNA 4930521123 gene"
		/protein_id="AA422549.1"
		/db_xref="GI:18490711"
		/translation="MAVNSHTENRRGALIPNGESLLKRSFNYELSPQRESGNVPS GRKQTLFTSYRVIPTTSGISIDPMLSPMPPDLMTNLTVEQVPFAANTIKGTIOAA PYGSGGQALPTKIVFNGCAIIRPAOLMTKASAAAGRPRLRTLDNWSNGIYVITCK GICAIHRCLVQLSMERPHQDMRPHLPDTFPHIQDMRPNP"
BASE COUNT		709 a 477 c 468 g 612 t
ORIGIN		

Query Match	77.0%	Score 771.2	DB 9	Length 2266
Best Local Similarity	98.9%	Prod. No. 8.2e-197		
Matches 787	Conservative 0	Mismatches 8	Indels 1	Gaps 1
QY	1	ATGCCATTGATCTGATGACGAACCTCAGTGTGGAACAACAGATATTGCTGCAACTTC	60	
Db	254	ATGCCATTGATCTGATGACGAACCTCAGTGTGGAACAACAGATATTGCTGCAACTTC	313	
QY	61	ATTAAAGGAACATATTCAGGCACTCCATATGTTGGCTGGGAAGACAAAGCTACTTTTAAA	120	
Db	314	ATTAAAGGAACATATTCAGGCACTCCATATGTTGGCTGGGAAGACAAAGCTACTTTTAAA	373	
QY	121	TTTACCTTCAGAAATGAGATGCCATTGAAATTTGGCCAGTTGATGTTGAAAGCTGCTCT	180	

```

Db      |||||
374  TTAGCTTCAGAAATGAGAGTCCATGTAATTCGCCCATGTAATGTAAGCTGCTCT 433
Oy      181  GCGTTGACCCGAGATTCCTCACTTGAACCTTAATATGACTGTTGAGTCTATGGAAT 240
Db      434  GCTGCGCCCGAGATTCCTCACTTGAACCTTAATATGACTGTTGAGTCTATGGAAT 493
Oy      241  TATGTAATTAATCTGGGAGAGGAATATGTCATCCACAGATGCTTTTCACTTATGTC 300
Db      494  TATGTAATTAATCTGGGAGAGGAATATGTCATCCACAGATGCTTTTCACTTATGTC 552
Oy      301  TATGGGCCCCCAGCTGACAGATATGAGGCCCACTCCCGATACGAGACCCCACTTGA 360
Db      553  TATGAGAGCCCACTGACAGATATGAGGCCCACTCCCGATACGAGACCCCACTTGA 612
Oy      361  GATATGAGCCCACTGAGGAATGAGAGCCCGCTGTGGGATACAGAGCTCACT 420
Db      613  GATATGAGCCCACTGAGGAATGAGAGCCCGCTGTGGGATACAGAGCTCACT 672
Oy      421  GTGCGATATGAGCCCACTCTTGATACGAGGCCCACTGACAGATATGAGGCCCA 480
Db      673  GTGCGATATGAGCCCACTCTTGATACGAGGCCCACTGACAGATATGAGGCCCA 732
Oy      481  CCTTGAATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGATATGAGACC 540
Db      733  CCTTGAATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGATATGAGACC 792
Oy      541  CCACTCTGAGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGCTGCGGATACAG 600
Db      793  CCACTCTGAGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGCTGCGGATACAG 852
Oy      601  GCCTCACTGCTGATCAAGAGCCAGCTCTCAAGATCTACAGAGCCAGCTCTCTGA 660
Db      853  GCCTCACTGCTGATCAAGAGCCAGCTCTCAAGATCTACAGAGCCAGCTCTCTGA 912
Oy      661  AACGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGA 720
Db      913  AACGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGA 972
Oy      721  TAAACCTTGAAGCTCAGCAAGCAAGAGGATCCCTAAATGAGATCAGATTAAGAGG 780
Db      973  TAAACCTTGAAGCTCAGCAAGCAAGAGGATCCCTAAATGAGATCAGATTAAGAGG 1032
Oy      781  ACGACTCAGCTTGAAG 796
Db      1033  ACGACTCAGCTTGAAG 1048

```

RESULT 4
HS250D10
LOCUS
DEFINITION
220895 bp. DNA. linear. PRI 05-JUN-2003
Human DNA sequence from clone CTA-250D10 on chromosome 22. Contains the genes for SREBP2 (sterol regulatory element binding transcription factor 2), NAGA (alpha-N-acetylgalactosaminidase), a gene similar to neuronal-specific septin 3, a pseudogene similar to ANT2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a genomic marker D2S1178, a CA repeat polymorphism, ESTs and a Cpg island, complete sequence.
299716
299716.4 GI:4456457
HTG; ANT2; Cpg Island; D2S1178; NAGA; septin 3; SREBP2.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk; Clone requests: clonerequests@sanger.ac.uk
On Mar 21, 1999 this sequence version replaced gi:416439.
During sequence assembly data is compared from overlapping clones.

FEATURES

source

```

1..220895
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="22"
/clone="CTA-250D10"
/location="11b-CIT978SK-A1"
17..29
/clone="11b-CIT978SK-A1"
17..29
/feature="2.6 copies 5 mer ATTTG 26% conserved"
33..325
/feature="1.301 of consensus"
326..338
/feature="1.301 of consensus"
326..338
/feature="1.301 of consensus"
667..683
/feature="2.1 copies 8 mer AGAAGACC 34% conserved"
740..751
/feature="2.1 copies 8 mer AGAAGACC 34% conserved"
1443..1457
/feature="2.4 copies 5 mer TTTTA 24% conserved"
1715..1997
/feature="7.5 copies 2 mer AC 21% conserved"
2103..2378
/feature="1.2378 of consensus"
2103..2378
/feature="1.2378 of consensus"
2844..2855
/feature="1.2855 of consensus"
2844..2855
/feature="1.2855 of consensus"
3746..3837
/feature="1.3837 of consensus"
3746..3837
/feature="1.3837 of consensus"
3838..4142
/feature="1.4142 of consensus"
3838..4142
/feature="1.4142 of consensus"
4143..4441
/feature="1.4441 of consensus"
4143..4441
/feature="1.4441 of consensus"

```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WormBase; Information on the WormBase database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormbase CTA-250D10 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.

VECTOR: pBAC108L

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

This sequence is the entire insert of clone CTA-250D10. The true left end of clone RPI-18601 is at 129979 in this sequence. The true left end of clone RPI-359316 is at 1339 in this sequence. The true right end of clone RPI-821D11 is at 23458 in this sequence. The true right end of clone RPI-359316 is at 118711 in this sequence.

LOCUS	AX359658	1413 bp	DNA	linear	PAT 13-FEB-2002
DEFINITION	AX359658	Sequence 4 from Patent WO0190185.			
ACCESSION	AX359658				
VERSION	AX359658.1	GI:18675409			
KEYWORDS					
SOURCE	unidentified				
ORGANISM	unidentified				
REFERENCE	1				
AUTHORS	Okamoto, R. and Sutovsky, P.				
TITLE	Plc3 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses thereof				
JOURNAL	Patent: WO 0190185-A 4 29-NOV-2001; QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES UNIVERSITY (US)				
FEATURES					
source	Location/Qualifiers				
	1..1413				
	/organism="unidentified"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32644"				
	/note="mammalia"				
	36..977				

Query Match	43.5%	Score 435.4	DB 6	Length 1413	
Best Local Similarity	74.7%	Pred. No. 2.9e-106			
Matches 609	Conservative	0	Mismatches 181	Indels 25	Gaps 4
QY	1	ATGCCATTGATCTGATGACGAACCTCACTGTTGAACAACCAAGTATTTGCTGCACAACTTC	60		
Db	258	ATGCCGTTGGCTGATGATGATGACCTGACCACTTGAACAACAAATTTTGGCCCCCAATAC	317		
QY	61	ATTAAAGGAACATATTCAGGACAGCTCCATATGGTGTGCTGGAAAGACAAAGCTACTTTTAA	120		
Db	318	ATTAAAGGAACCATTCAGGACAGCTCCAGGTGTGTGCTGGAAAGACAAAGCTGTTTAAAG	377		
QY	121	TTAGTCTTCAGAAATGAGATGATGCCATTGAATTTGGCCAGTGTGATGTAAGAACTGCTCT	180		
Db	378	TTATCTCTTCAGAAAGAGATGATGCCATTGCAATTTGCCCCAAGTATGTAAGAACTGCTCT	437		
QY	181	GCTGTGGCCGAGATTTCCACTTAGAACCTTAAATGACTGTGACGCTTATGAGAAATT	240		
Db	438	GCTGTGCGAAGGAATTCACCTTGAAAGTGAATTAATCTGTTTCGACACTTCACAGACTG	497		
QY	241	TATGTAATTACTGAGGAAGGAAT---ATGTCACTCCACAGATGCTTTGTTCAG-----	292		
Db	498	TACATAAATTACTGTCCCAAGGGGCTCAAGTGTGCTCTTCAACAACAACCTTGTTCACACATAT	557		
QY	293	----TTATGTGCTATGAGGGGCCCCCACTGACAGATATGAGGCCCACTCCCGGATACGGA	348		
Db	558	CCAAATGTGATTTATGAGACCCCCCAACAAGATATTAAGTTCACCAACAGGGGAATATGGA	617		
QY	349	GCCCCACCTGCAGATATGAGACCCCAACCCGTAAGAAAATGAAGGCCCGCTGTGTGGAATAC	408		
Db	618	ACTCACCAAGAAATATGAGACCCCAACAAGGGGAGTATGAGGCCCACTATGGAATAT	677		
QY	409	AGAGCTCACTGTGCGATATGAGACCCCACTCTTGTGATACGAGAGCCCACTGTGACAGA	468		
Db	678	GGAGGCCCGCTGTGGAATATGAAGTCCCACTGGGGGATATGAGAGTCCCACTGTGGGGA	737		

Qy	469	TATGAGCCCCACCTCTAGAGATATGAGCCCCACCTCTGATATGAGACCCACCTCTC	528
Db	738	TATGAGATCCCACTGGGGGATATGAGCCCCACCTGGGGGATATGAGATCCCACTGGG	797
Qy	529	GGATATGAGCCCCACCTCTGGAGATATGAGCCCCACCTGACAGAAATGAAGCCGCT	588
Db	798	GGATATGATGCCCACTGGGGGATATGAGCCCCACCTGACAGATATGAGCCCCACCA	857
Qy	589	GCGGATACAGAGCTTCACTGTGATCAGAGAGCCAGGCTCAGAAATTCACAGCAGCC	648
Db	858	GCTGGAATGAGGCCCTACCCCTGCATATGAAGCTTCATCTGTGAAATACAGCTGCC	917
Qy	649	-----CAGGCTCTGAAAAGAGGCTTCTTCCCTCTGCTCTCTCTTCTCAG	696
Db	918	TCTCAGAGATCTATGACAGCTCAGCAGAGACCTTCTTCCCACTACCTATCTTTCTAG	977
Qy	697	GTCATCTTAACCTTCTAGATGTAAACCTGTAAGCTCACCAAGCAAGAGGTACCT	756
Db	978	GTCATTTACACCTTCTCAGAGTTAACTTTAGAGCTCACCAAGCAAG-GGACCTT	1036
Qy	757	AAATTTGAGTCAGATATGAGAGCACTCAGCT	791
Db	1037	AAATCTGATCAGATATGAGAGCACTCAGCT	1071

	RESULT 6	
LOCUS	BX296515	
DEFINITION	BX296515 179222 bp DNA linear HTG 26-MAY-2003 Sus scrofa clone pIge-12ID21, *** SEQUENCING IN PROGRESS ***, 6 unordered pieces.	
ACCESSION	BX296515	
VERSION	BX296515.6 GI:31076160	
KEYWORDS	HTG; HTGS PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULTOP. Sus scrofa (pig)	
SOURCE	Sus scrofa	
ORGANISM	Bufo boreas; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.	
REFERENCE	Tracey, A. Direct Submission Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On May 26, 2003 this sequence version replaced gi:31043704.	
AUTHORS	Genome Center	
TITLE	Center: Wellcome Trust Sanger Institute	
JOURNAL	Center code: SC Web site: http://www.sanger.ac.uk Contact: humquerry@sanger.ac.uk Project information Center project name: BR12ID21	
COMMENT	----- Summary Statistics ----- Assembly program: XGAP4; version 4.5 Chemistry: Dye-terminator; 10% of reads Consensus quality: 177502 bases at least Q40 Consensus quality: 177971 bases at least Q30 Consensus quality: 178283 bases at least Q20 Insert size: 178722; sum-of-contigs Insert size: 167315; 12.1% error; agarose-fp Quality coverage: 7.82x in Q20 bases; sum-of-contigs Quality coverage: 8.59x in Q20 bases; agarose-fp ----- NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 5850: contig of 5850 bp in length * 5851 5950: gap of 100 bp * 5951 74264: contig of 68314 bp in length	

[illegible]

```

DB              55994  AGGACTCAGGTATGTG 55993

RESULT 7
LOCUS      BX470149
DEFINITION Danio rerio clone DKEX-242K7, *** SEQUENCING IN PROGRESS ***, 50
ACCESSION  BX470149
VERSION    BX470149.3  GI:30424228
HTG; HTGS PHASBI.
SOURCE     Danio rerio (zebrafish)
ORGANISM   Danio rerio
            Bkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
            1 (bases 1 to 22469)
            Burton.J.
REFERENCE   Direct Submission
            Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zf1sh-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
            On May 7, 2003 this sequence version replaced gi:30387077.
AUTHORS    Genome Center
TITLE      Center: Wellcome Trust Sanger Institute
JOURNAL    Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zf1sh-help@sanger.ac.uk
            Project Information
            -----
            Center project name: ZK242K7
            -----
            Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 197958 bases at least Q40
            Consensus quality: 206715 bases at least Q30
            Consensus quality: 212351 bases at least Q20
            Insert size: 217569; sum-of-contigs
            Insert size: 165837; 6.2% error; agarose-fp
            Quality coverage: 2.48x in Q20 bases; sum-of-contigs Quality
            coverage: 3.95x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 50 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be prefiltered.
            *
            *
            1      6656: contig of 6656 bp in length
            *      6657      6756: gap of 100 bp
            *      6757      10255: contig of 3499 bp in length
            *      10256      10356: gap of 100 bp
            *      10356      13223: contig of 2868 bp in length
            *      13224      13323: gap of 100 bp
            *      13324      19935: contig of 6612 bp in length
            *      19936      20035: gap of 100 bp
            *      20036      22597: contig of 2562 bp in length
            *      22598      22697: gap of 100 bp
            *      22698      28563: contig of 5872 bp in length
            *      28570      28669: gap of 100 bp
            *      28670      37965: contig of 9296 bp in length
            *      37966      38065: gap of 100 bp
            *      38066      41140: contig of 3075 bp in length
            *      41141      41240: gap of 100 bp
            *      41241      45108: contig of 3868 bp in length
            *      45109      45208: gap of 100 bp
            *      45209      47730: contig of 2522 bp in length
            *      47731      47830: gap of 100 bp
            *      47831      51984: contig of 4054 bp in length
            *      51984      51985: gap of 100 bp
            *      51985      55857: contig of 3873 bp in length

```

```

* 55858 55957: gap of 100 bp
* 55958 60498: contig of 4541 bp in length
* 60499 60598: gap of 100 bp
* 60599 63641: contig of 3043 bp in length
* 63642 63741: gap of 100 bp
* 63742 72674: contig of 8933 bp in length
* 72675 72774: gap of 100 bp
* 72775 75502: contig of 2838 bp in length
* 75503 75702: gap of 100 bp
* 75703 78063: contig of 2360 bp in length
* 78063 78162: gap of 100 bp
* 78163 84485: contig of 6323 bp in length
* 84486 84585: gap of 100 bp
* 84586 87553: contig of 296 bp in length
* 87554 87653: gap of 100 bp
* 87654 92487: contig of 4834 bp in length
* 92488 92587: gap of 100 bp
* 92588 96104: contig of 3517 bp in length
* 96105 96204: gap of 100 bp
* 96205 101479: contig of 5275 bp in length
* 101480 101579: gap of 100 bp
* 101580 104979: contig of 3400 bp in length
* 104980 105079: gap of 100 bp
* 105803 108003: contig of 2924 bp in length
* 108004 108103: gap of 100 bp
* 108104 113151: contig of 5048 bp in length
* 113152 113251: gap of 100 bp
* 113252 116585: contig of 3334 bp in length
* 116586 116685: gap of 100 bp
* 116686 119346: contig of 2661 bp in length
* 119347 119446: gap of 100 bp
* 119447 123125: contig of 3679 bp in length
* 123126 131266: contig of 8041 bp in length
* 131267 131366: gap of 100 bp
* 131367 136193: contig of 4827 bp in length
* 136194 136293: gap of 100 bp
* 136294 139482: contig of 3189 bp in length
* 139483 139582: gap of 100 bp
* 139583 141875: contig of 2293 bp in length
* 141876 141975: gap of 100 bp
* 141976 144122: contig of 2147 bp in length
* 144123 144222: gap of 100 bp
* 144223 146968: contig of 2746 bp in length
* 146969 147068: gap of 100 bp
* 147069 149399: contig of 2331 bp in length
* 149400 149499: gap of 100 bp
* 149500 153963: contig of 4464 bp in length
* 153964 154063: gap of 100 bp
* 154064 156579: contig of 2516 bp in length
* 156580 156679: gap of 100 bp
* 156680 158928: contig of 2249 bp in length
* 158929 159028: gap of 100 bp
* 159029 164542: contig of 5514 bp in length
* 164543 164642: gap of 100 bp
* 164643 172054: contig of 7412 bp in length
* 172055 172154: gap of 100 bp
* 172155 175210: contig of 3056 bp in length
* 175211 175310: gap of 100 bp
* 175311 179379: contig of 4069 bp in length
* 179380 179479: gap of 100 bp
* 179480 182698: contig of 3219 bp in length
* 182699 182798: gap of 100 bp
* 182799 186779: contig of 3981 bp in length
* 186780 186879: gap of 100 bp
* 186880 195626: contig of 8747 bp in length
* 195627 195726: gap of 100 bp
* 195727 199396: contig of 3670 bp in length
* 199397 199496: gap of 100 bp
* 199497 206412: contig of 6916 bp in length
* 206413 209619: gap of 100 bp
* 209620 209719: gap of 100 bp

```

```

FEATURES
    source                * 209720 219083: contig of 9364 bp in length
                        * 219084 219183: gap of 100 bp
                        * 219184 222469: contig of 3286 bp in length.
                        Location/Qualifiers
                            1..222469
                                /organism="Danio rerio"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:7955"
                                /clone="DKRY-242K"
                                /clone_1b="DanioKey"
                                1..6656
                                    /note="assembly_fragment:01359
                                        fragment_chain:1"
                                6757..10255
                                    /note="assembly_fragment:01385
                                        fragment_chain:1"
                                10356..13223
                                    /note="assembly_fragment:00592
                                        fragment_chain:1"
                                13374..19935
                                    /note="assembly_fragment:00237
                                        fragment_chain:1"
                                20036..22597
                                    /note="assembly_fragment:00949
                                        fragment_chain:1"
                                22698..28569
                                    /note="assembly_fragment:00615
                                        fragment_chain:1"
                                28670..37965
                                    /note="assembly_fragment:01306
                                        fragment_chain:2"
                                38066..41140
                                    /note="assembly_fragment:00487
                                        fragment_chain:2"
                                41241..45108
                                    /note="assembly_fragment:00232
                                        fragment_chain:2"
                                45209..47730
                                    /note="assembly_fragment:00935
                                        fragment_chain:2"
                                47831..51884
                                    /note="assembly_fragment:01415
                                        fragment_chain:2"
                                51985..55857
                                    /note="assembly_fragment:00878
                                        fragment_chain:3"
                                55958..60498
                                    /note="assembly_fragment:01208
                                        fragment_chain:3"
                                60599..63641
                                    /note="assembly_fragment:01266
                                        fragment_chain:3"
                                63742..72674
                                    /note="assembly_fragment:01329
                                        fragment_chain:3"
                                72775..75502
                                    /note="assembly_fragment:01183
                                        fragment_chain:4"
                                75703..78062
                                    /note="assembly_fragment:01179
                                        fragment_chain:4"
Query Match      27.3%; Score 273.6; DB 2; Length 222469;
Beet Local Similarity 75.0%; Pred. No. 2.8e-62;
Matches 372; Conservative 0; Mismatches 114; Indels 10; Gaps 2;
QY      301 TATGGAGCCCACTGAGATATGAGCCCACTCCCGATATGAGCCCACTCA 360
DB      197080 TACGAGCCCACTGAGAAAGCTGAGACCCCAAGAGATATGAGCCCACTCG 197139
QY      361 GGAATGAGCCCAACCCGTAGAAATGAGCCCGCTGTGGATACAGAGCTCACT 420
DB      197140 GGAATGAGAGCTCACTGAGATATGAGCCCGCTCCAGATATGAGAGCTCACT 197199

```


Location/Qualifiers
1. .168425

	a	c	g	t	others
BASE COUNT	44391	39572	38747	44015	1700
ORIGIN					

303 TGGGGCCCACTGCAGGATATGGAGCCCCCACCCTCCCGGATACGGAGCCCCCACCCTGCAGG 362

Db 66773 CACTTAGGT 66781

sequence.

ORGANIS

AUTHORS

JOURNAL

JOURNAL

JOURNAL

JOURNAL

COMMENT

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
 On May 30, 2003 this sequence version replaced gi:29124187.
 ----- Genome Center
 Center: Department Of Chemistry And Biochemistry
 The University Of Oklahoma
 Center code:UOKNOR

FEATURES

Location/Qualifiers
 source 1. 224086
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /map="15"
 /clone="ip23-204m3"
 /clone_id="RPCT - 23 Female (C57BL/6J) Mouse BAC Library"

BASE COUNT 55592 a 55124 c 56095 g 57275 t
 ORIGIN

Query Match 23.1% Score 231.4; DB 10; Length 224086;
 Best Local Similarity 67.1%; Pred. No. 6.8e-51;
 Matches 328; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 303 TGGGGCCCACTGTCAGAGATATGAGGCCCACTCCGGATACGAGGCCCACTTGAGG 362
 DB 162839 TGGATATGAGAGCTGATGATATGCTCCCTCCCTCTATATATGATCTACCCCAATGGG 162898
 QY 363 ATATGAGAGCCCACTGATGATATGAGGCCCGCTGATGATACAGAGCTTCACTGT 422
 DB 162899 CTATGAGATTCACCTCTGATATGAGCCCTCCCTGATGATACGATCTCCCACTCC 162958
 QY 423 GCGATATGAGAGCCCACTCTGATATGAGGCCCGCTGATGATATGAGGCCCACT 482
 DB 162959 TGGATATGAGAGCCCACTATGATATGAGGCCCACTCCGATACGATACCACT 163018
 QY 483 TCTAGATATGAGAGCCCACTCTGATATGAGGCCCACTCTGATATGAGGCCCG 542
 DB 163019 TATGAGATCCGGTTCCTCCCACTCCCAATATGAGGCCCGCTATGAGGATGCAATCC 163078
 QY 543 ACCTCTGATATGAGAGCCCACTCTGATATGAGGCCCGCTGATATGAGGCCCG 602
 DB 163079 TCTTTCAGAGGCGTATGATATCTCTGATATGAGGCCCACTCTGATACAGAGG 163138
 QY 603 CTACCTGCTGATATGAGAGGCCCGCTGATATGAGGCCCACTCTGATATGAGGCCCG 662
 DB 163139 TCAACCTGCTGAGGCTGATATGAGGCCCGCTGATATGAGGCCCGCTGATAT 163198
 QY 663 CGAGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722
 DB 163199 CGAGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 163258
 QY 723 AACCTTGAAGACTGACCAAGCAAGAGTACCTTAATGTAAGTCAAGATTAAGAGGAC 782
 DB 163259 AACCTTGAAGTTCACCAAGCAAGCAAGAGTACCTTAATGTAAGTCAAGATTAAGAGGAC 163318
 QY 783 GACTCAGCT 791
 DB 163319 CACTTAGGT 163327

RESULT 10
 AC107527/c 253149 bp DNA linear HTG 13-MAY-2003
 LOCUS Rattus norvegicus clone CH230-92M24, WORKING DRAFT SEQUENCE, 3
 DEFINITION Unordered pieces.
 AC107527
 ACCESSION AC107527.5 GI:30580771
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE

1 (bases 1 to 253149)
 Ratius.
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Arguano, D.,
 Anyalobechi, V., Ayodeji, A., Ayodeji, M., Baca, B., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benamed, F.,
 Bielwalo, K., Blair, J., Blankensbury, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
 Chaco, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Crease, A., D'Souza, L.,
 Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dedrich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Bayes, K.,
 Egan, A., Becotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
 Hollins, B., Howells, S., Huylk, S., Hume, J., Illedit, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenz, L., Louie, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
 Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, B.,
 Milosavljevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munitas, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwaokemele, O., Okunolu, G., Olamugbo, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sores, J.,
 Steele, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingley, A., Trejos, Z., Umani, K.,
 Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G., and Gibbs, R.A.

COMMENT

REFERENCE
 TITLES
 AUTHORS
 JOURNAL
 REFERENCE
 TITLES
 AUTHORS
 JOURNAL
 COMMENT
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 13, 2003 this sequence version replaced gi:23664681.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/atl/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence

ORGANISM Sus scrofa
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 REFERENCE 1 (bases 1 to 129624)
 AUTHORS Burton, J.
 TITLE Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Apr 2, 2003 this sequence version replaced gi:29335441.
 COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 Project Information
 Center project name: b7231K18
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 113740 bases at least Q40
 Consensus quality: 118187 bases at least Q30
 Consensus quality: 120712 bases at least Q20
 Insert size: 126024; sum-of-contigs
 Insert coverage: 154528; 4.6% error; agarose-fp
 Quality coverage: 2.35x in Q20 bases; sum-of-contigs Quality
 coverage: 3.19x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 37 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1 4262: contig of 4262 bp in length
 * 4263 4362: gap of 100 bp
 * 4363 6991: contig of 2629 bp in length
 * 6992 7091: gap of 100 bp
 * 7092 9480: contig of 2389 bp in length
 * 9481 9580: gap of 100 bp
 * 9581 11688: contig of 2108 bp in length
 * 11689 11788: gap of 100 bp
 * 11789 14658: contig of 2870 bp in length
 * 14659 14758: gap of 100 bp
 * 14759 18440: contig of 3682 bp in length
 * 18441 18540: gap of 100 bp
 * 18541 20818: contig of 2278 bp in length
 * 20819 20918: gap of 100 bp
 * 20919 22969: contig of 2051 bp in length
 * 22970 23069: gap of 100 bp
 * 23070 27584: contig of 4515 bp in length
 * 27585 27684: gap of 100 bp
 * 27685 33211: contig of 5527 bp in length
 * 33212 33311: gap of 100 bp
 * 33312 35521: contig of 2210 bp in length
 * 35522 35621: gap of 100 bp
 * 35622 38528: contig of 2907 bp in length
 * 38529 38628: gap of 100 bp
 * 38629 42063: contig of 3435 bp in length
 * 42064 42163: gap of 100 bp
 * 42164 44164: contig of 2001 bp in length
 * 44165 44264: gap of 100 bp
 * 44265 47707: contig of 3443 bp in length
 * 47708 47807: gap of 100 bp
 * 47808 50737: contig of 2930 bp in length
 * 50738 50837: gap of 100 bp
 * 50838 57066: contig of 6229 bp in length
 * 57067 57166: gap of 100 bp
 * 57167 59978: contig of 2812 bp in length
 * 59979 60078: gap of 100 bp
 * 60079 62560: contig of 2482 bp in length

FEATURES
 source * 62561 62660: gap of 100 bp
 * 62661 64963: contig of 2303 bp in length
 * 64964 65063: gap of 100 bp
 * 65064 67191: contig of 2128 bp in length
 * 67192 67292: gap of 100 bp
 * 67292 72005: contig of 4714 bp in length
 * 72006 72105: gap of 100 bp
 * 72106 76595: contig of 4490 bp in length
 * 76596 76696: gap of 100 bp
 * 76696 79826: contig of 3130 bp in length
 * 79826 79926: gap of 100 bp
 * 79926 83540: contig of 3615 bp in length
 * 83541 83640: gap of 100 bp
 * 83641 86276: contig of 2636 bp in length
 * 86277 86376: gap of 100 bp
 * 86377 94620: contig of 8244 bp in length
 * 94621 94720: gap of 100 bp
 * 94721 97520: contig of 2800 bp in length
 * 97521 97620: gap of 100 bp
 * 97621 99784: contig of 2164 bp in length
 * 99785 99884: gap of 100 bp
 * 99885 102112: contig of 2228 bp in length
 * 102113 102212: gap of 100 bp
 * 102213 104242: contig of 2030 bp in length
 * 104243 104342: gap of 100 bp
 * 104343 110710: contig of 6368 bp in length
 * 110711 110810: gap of 100 bp
 * 110811 11367: contig of 2557 bp in length
 * 11368 11367: gap of 100 bp
 * 11368 11367: gap of 100 bp
 * 11368 11505: contig of 2038 bp in length
 * 11506 11505: gap of 100 bp
 * 11506 121874: contig of 6269 bp in length
 * 121875 121974: gap of 100 bp
 * 121975 127096: contig of 5122 bp in length
 * 127097 127196: gap of 100 bp
 * 127197 129624: contig of 2428 bp in length.
 Location/Qualifiers
 1..129624
 /organism="Sus scrofa"
 /mol_type="genomic DNA"
 /db_xref="taxon:9823"
 /clone="P1g8-231K18"
 /clone_1b="P1g8"
 1..4262
 /note="assembly fragment:00361
 fragment_chain:1"
 4363..6991
 /note="assembly fragment:00108
 fragment_chain:1"
 7092..9480
 /note="assembly fragment:00460
 fragment_chain:1"
 9581..11688
 /note="assembly fragment:00959
 fragment_chain:1"
 11789..14658
 /note="assembly fragment:00995
 fragment_chain:2"
 14759..18440
 /note="assembly fragment:00608
 fragment_chain:2"
 18541..20818
 /note="assembly fragment:01101
 fragment_chain:2"
 20919..22969
 /note="assembly fragment:00996
 fragment_chain:2"
 23070..27584
 /note="assembly fragment:00217
 fragment_chain:3"
 27685..33211
 /note="assembly fragment:00019
 fragment_chain:3"

Query	DB	Matches	263j	Conservative	0j	Mismatches	91j	Indels	0j	Gaps	0j
Query Match		20.84j	Score 208.4j	DB 2j	Length 129624j						
Best Local Similarity		74.3j	Pred. No. 9.9e-45j								
misc_feature		33312..35521	/note="assembly_fragment:00548								
fragment_chain:3"											
misc_feature		35622..38528	/note="assembly_fragment:00815								
fragment_chain:4"											
misc_feature		38629..42063	/note="assembly_fragment:00246								
fragment_chain:4"											
misc_feature		42164..44164	/note="assembly_fragment:00859								
fragment_chain:5"											
misc_feature		44265..47707	/note="assembly_fragment:00964								
fragment_chain:5"											
misc_feature		47808..50737	/note="assembly_fragment:01192								
fragment_chain:6"											
misc_feature		50838..57066	/note="assembly_fragment:00542								
fragment_chain:6"											
misc_feature		57167..59978	/note="assembly_fragment:01304								
fragment_chain:7"											
misc_feature		60079..62560	/note="assembly_fragment:00948								
fragment_chain:7"											
misc_feature		62661..64963	/note="assembly_fragment:00016"								
fragment_chain:7"											
misc_feature		65064..67193	/note="assembly_fragment:00065"								
fragment_chain:7"											
misc_feature		67292..72005	/note="assembly_fragment:00070"								
fragment_chain:7"											
misc_feature		72106..76595	/note="assembly_fragment:00136"								
fragment_chain:7"											
misc_feature		76596..79825	/note="assembly_fragment:00162"								
fragment_chain:7"											
misc_feature		79926..83540	/note="assembly_fragment:00367"								
fragment_chain:7"											
misc_feature		83641..86276	/note="assembly_fragment:00492"								
fragment_chain:7"											
misc_feature		86377..94620	/note="assembly_fragment:00545"								
fragment_chain:7"											
misc_feature		94721..97520	/note="assembly_fragment:00745"								
fragment_chain:7"											
misc_feature		97621..99784	/note="assembly_fragment:00801"								
fragment_chain:7"											
Query Match		20.84j	Score 208.4j	DB 2j	Length 129624j						
Best Local Similarity		74.3j	Pred. No. 9.9e-45j								
Matches	263j	Conservative	0j	Mismatches	91j	Indels	0j	Gaps	0j		
Query	296	TTGTCTATGGAGCCCCCACTGACGATATGAGGCCCACTCCCGATACGAGCCCCAC	355								
DB	65475	TAGGATATGAGAGGCCAACCAGAGATATGAGACCCCACTCCAGATACGAGCCCCAC	65416								
Query	356	CTGAGAGATATGAGAGGCCCAACCGTATGAGAAATGAAGCCCGCTGTGGATACAGAGCCT	415								
DB	65415	CTCCAGAGATACGAGGCCCACT									

Db	65175	ATGAGGCCCACTGCAAGATATGATGCCCCGCTCAGATATGAGCTCAC	65122
RESULT 13			
LOCUS	AC020698	148418 bp	DNA linear PRI 01-MAR-2002
DEFINITION	Homo sapiens BAC clone R11-45F23 from 4, complete sequence.		
ACCESSION	AC020698		
VERSION	AC020698.4	GI:11120934	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 148418)		
AUTHORS	Suleston, J.R. and Waterston, R.		
TITLE	Toward a complete human genome sequence		
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)		
MEDLINE	98063792		
PubMed	9847074		
REFERENCE	2 (bases 1 to 148418)		
AUTHORS	Nguyen, C., Drone, K., Hawkins, M. and Ureta, M.		
TITLE	The sequence of Homo sapiens BAC clone R11-45F23		
JOURNAL	Unpublished (2001)		
REFERENCE	3 (bases 1 to 148418)		
AUTHORS	Waterston, R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
REFERENCE	4 (bases 1 to 148418)		
AUTHORS	Waterston, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
REFERENCE	5 (bases 1 to 148418)		
AUTHORS	Waterston, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
REFERENCE	6 (bases 1 to 148418)		
AUTHORS	Waterston, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (21-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
REFERENCE	7 (bases 1 to 148418)		
AUTHORS	Waterston, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
COMMENT	On Nov 8, 2000 this sequence version replaced gi:7630812.		

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

chemistry, or covered by high quality data (i.e., phred quality >=

as compressions and repeats; all regions were covered by sequence

from more than one source; and the assembly was confirmed by restriction digest.


```

OY 427 TATGAGAGCCCACTCTTGATACGAGAGCCCACTGAGATATGAGCCCACTCTA 486
DB 43181 TGTCTAGCTCCACCTCCAGATGCTTACTCTCAACCCCAAGATGCTTACTCTCA 43240
OY 487 GGATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGATATGAGAGCCCACT 546
DB 43241 GGATCTTACTGCTTCAACCCCAAGATGCTTACTCTCAACCCCAAGATGCTTACTCTCA 43300
OY 547 CTGATATGAGAGCCCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 606
DB 43301 CAGAGCTGCTTACTGCTTCAACCCCAAGATGCTTACTCTCAACCCCAAGATGCTTACTCTCA 43360
OY 607 CCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 649
DB 43361 CCCCAGATGCTTACTGCTTCAACCCCAAGATGCTTACTCTCAAC 43403

RESULT 14
AC103031/c
LOCUS AC103031 235381 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-203117, WORKING DRAFT SEQUENCE, 3
unordered pieces.
ACCESSION AC103031
VERSION AC103031.5 GI:30580672
KEYWORDS HTG: HTGS PHASRI; HTGS DRAFT; HTGS_FUTLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 235381)
Munzy,D.Marie, Metzker,M.Lee, Adrianzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Barnmed,F.,
Bilwalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesari,H., Cente,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Drepper,H., Dugan-Rocha,S., Dunn,A., Durkin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Friser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgievski,B., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M.,
Gunnarsson,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howell,S., Huik,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kovis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,Y., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshuewa,L., Louisedge,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangun,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
Milosavljevic,A., Miner,G., Mijic,B., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L.,
Nakervic,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwoketeme,O., Okonou,G., Olarnunagoon,A., Pal,S., Parks,K.,
Paternek,S., Paul,H., Perez,A., Perez,L., Pfankech,C.,
Plopper,F., Polidexter,A., Popovic,D., Primus,B., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojao,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorrelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Thorpe,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Uemami,K.,

```

COMMENT

REFERENCE

Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,
 Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,K., Zhao,S., Zhou,D., von
 Wernershausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 and Gibbs,R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 235381)
 Worley,K.C.
 Direct Submission
 Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 235381)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 13, 2003 this sequence version replaced gi:23123692.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the end of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

COMMENT

REFERENCE

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GIZG
 Center clone name: CH230-203117
 Summary Statistics
 Assembly program: Atlas 3.0
 Consensus quality: 221482 bases at least Q40
 Consensus quality: 223612 bases at least Q30
 Consensus quality: 224834 bases at least Q20
 Baited insert size: 231669; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

FEATURES

SOURCE

1. 235381
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-203117"
 1. 1102
 /note="wgs_end_extension"

misc_feature

```

misc_feature      clone_end:Sp6"
                  4245_4594
                  /note="clone boundary
                  clone_end:Sp6
                  site:BCORI
misc_feature      end sequence:BH364165"
                  complement(230311..230677)
                  /note="clone boundary
                  clone_end:T7
                  site:BCORI
misc_feature      end sequence:BH364164"
                  231434..232989
                  /note="xgs_end_extension
                  clone_end:T7"

BASE COUNT      5674 a 53524 c 55897 g 56470 t 9816 others
ORIGIN
Query Match      11.8%; Score 118.2; DB 2; Length 23581;
Best Local Similarity 54.2%; Pred. No.2,3e-20;
Matches 240; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

QY      293 TTATGTCTATGAGGCCCCCACTGCAGATATGAGGCCCACTCCCGATACGAGCCC 352
DB      141802 TTCTATTATACATGAAACCTTCTCTATTACATGAGACCTCTCTATTACATGAGCAC 141743
QY      353 CACCTGAGATATGAGCCCAACCCGTAGAAATGAGGCCCGCTGTGGATACAGAG 412
DB      141742 CTCCTTATTAATGAGAGCACTTCTCTATTACATGAGACCTTCTCTATTACATGAGAG 141683
QY      413 CCTCACCTGAGATATGAGGCCCACTCTGTGATATGAGGCCCACTGCAGATATG 472
DB      141682 CACCTCTTATTATACATGAGCACTTCTCTATTACATGAGACCTCTCTATTACATG 141623
QY      473 GAGCCCCACCTCTAGATATGAGCCCACTCTGTGATATGAGACCCCACTCTGTGAT 532
DB      141622 GATCACCTCTTATTATACATGAGCACTTCTCTATTACATGAGACCTTCTCTATTAC 141563
QY      533 ATGAGAGCCCACTCTGTGATATGAGGCCCACTCTGTGAGAAATGAGGCCCGCTGTGCG 592
DB      141562 ATGAGATCACCTCTTATTATACATGAGCACTTCTCTATTACATGATCACCTCTTTAT 141503
QY      593 GATACAGAGCCTCACCTGTGATATGAGGCCCACTCTGTGATATGAGAGCCCAAG 652
DB      141502 TATACATGAGCACTCTCTATTATACATGATACCTCTTTATTACATGAGCACTTCTC 141443
QY      653 CTCCTGAAAAGAGGCTTCTCTCCCTGCTCTCTCTCTCTGAGCTCATTTCAACTT 712
DB      141442 TATTACACAGAGCACTTCTCTATTACATGAGCACTCTCTATTACATGAGCACTT 141383
QY      713 CTAAGATGTAACCTTGAAGACT 735
DB      141382 CTCATTATTAATTAATGAGCACT 141360

RESULT 15
AC108000      167587 bp      DNA      linear      PRI 21-MAY-2002
LOCUS      Homo sapiens chromosome 15, clone CTD-2116G1, complete sequence.
DEFINITION      AC108000
ACCESSION      AC108000.5 GI:21039854
VERSION      HTG.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 167587)
AUTHORS      Birren,B., Linton,L., Nuebaum,C., Lander,E., Ali,A., Allen,N.,
TITLES      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
JOURNAL      Homo sapiens chromosome 15, clone CTD-2116G1
REFERENCE      2 (bases 1 to 167587)
AUTHORS      Birren,B., Linton,L., Nuebaum,C., Lander,E., Ali,A., Allen,N.,
TITLES      Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouhgalter,B.,
JOURNAL      Brown,A., Camarata,J., Campojiano,A., Chang,D., Chazaro,B.,

```

```

REFERENCE
AUTHORS      Cookel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
TITLES      Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
JOURNAL      Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N.,
AUTHORS      Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
TITLES      Kamat,A., Karatas,A., Kelle,C., Lacroque,K., Lamazares,R.,
JOURNAL      Lander,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
AUTHORS      Macdonald,P., Major,J., Marcuis,N., Matthews,C., McCarthy,M.,
TITLES      McEwan,P., McKernan,K., Naylor,T., Nguyen,C., Nicol,R., Norbu,C.,
JOURNAL      Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
AUTHORS      Peterson,K., Phunhkhang,P., Piere,N., Pollara,V., Raymond,C.,
TITLES      Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
JOURNAL      Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnpback,R., Seaman,S.,
AUTHORS      Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
TITLES      Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
JOURNAL      Topham,K., Travers,M., Travis,N., Trigilio,J., Vasilev,H.,
AUTHORS      Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G.,
TITLES      Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL      Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 167587)
REFERENCE
AUTHORS      Birren,B., Linton,L., Nuebaum,C., Lander,E., Ali,A., Allen,N.,
TITLES      Bouhgalter,B., Brown,A., Camarata,J., Campojiano,A., Chang,D.,
JOURNAL      Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
AUTHORS      Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
TITLES      Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
JOURNAL      Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N.,
AUTHORS      Hago,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
TITLES      Kamat,A., Karatas,A., Kelle,C., Lacroque,K., Lamazares,R.,
JOURNAL      Lander,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
AUTHORS      Macdonald,P., Major,J., Marcuis,N., Matthews,C., McCarthy,M.,
TITLES      McEwan,P., McKernan,K., Naylor,T., Nguyen,C., Nicol,R., Norbu,C.,
JOURNAL      Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
AUTHORS      Peterson,K., Phunhkhang,P., Piere,N., Pollara,V., Raymond,C.,
TITLES      Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
JOURNAL      Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnpback,R., Seaman,S.,
AUTHORS      Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
TITLES      Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
JOURNAL      Topham,K., Travers,M., Travis,N., Trigilio,J., Vasilev,H.,
AUTHORS      Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.D., Young,G.,
TITLES      Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL      Submitted (21-MAY-2002) Whitehead Institute/MIT Center for Genome

```

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 21, 2002 this sequence version replaced g1:20336146.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code

Web site: <http://www-beq.wi.mit.edu>

-----Project Information-----
 contact: sequence_submission@genome.wi.mit.edu

Project Information
Center project name: L24580

Center clone name: 2116 G 1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

FEATURES

Location/Qualifiers

Source

UNRES	
source	Location/Qualifiers
1. 167587	/clone="CTD-2116G1"
	/clone_1bp="CTD1 Human BAC"
	complement(11..144)
	/rpt_family="12"
	complement(145..356)
	/rpt_family="A1usg/x"
	complement(357..1109)
	/rpt_family="12"
	3713..3828
	/rpt_family="11MB"
	complement(3894..4006)
	/rpt_family="MIR"
	4574..4603
	/rpt_family="G(AAA)n"
	complement(4976..5167)
	/rpt_family="MIR"
	5639..5775
	/rpt_family="L1"
	6745..6766
	/rpt_family="AT_rich"
	complement(8290..8370)
	/rpt_family="MIR3"
	9077..9288
	/rpt_family="MIR"
	9496..10838
	/rpt_family="11MD1"
	10839..11162
	/rpt_family="A1usx"
	11163..13229
	/rpt_family="11MD1"
	complement(13118..13414)
	/rpt_family="A1uv"
	13667..14061
	/rpt_family="12"
	14366..14395
	/rpt_family="AT_rich"
	14748..15052
	/rpt_family="A1usx"
	complement(15603..15768)
	/rpt_family="MIR"
	complement(15878..15932)
	/rpt_family="12"
	complement(15988..16069)
	/rpt_family="MER81"
	complement(16309..16442)
	/rpt_family="MER46A"
	complement(16557..16730)
	/rpt_family="MIR"
	complement(16746..16830)
	/rpt_family="MIR"
	16845..16939
	/rpt_family="MER117"

```
repeat_region complement(17267. .17436)
/rpt_family="MIR3"
repeat_region complement(17914. .18014)
/rpt_family="MIR103"
repeat_region complement(19124. .19415)
/rpt_family="AluJo"
repeat_region complement(19571. .19636)
/rpt_family="L2"
repeat_region 19637. .19926
/rpt_family="AluJo"
repeat_region complement(19927. .19984)
/rpt_family="L2"
repeat_region complement(20449. .21788)
/rpt_family="L2"
repeat_region 21985. .22144
/rpt_family="MIR"
repeat_region 22726. .22838
/rpt_family="Char1ieg"
repeat_region 23014. .23044
/rpt_family="AT_rich"
repeat_region 23140. .23256
/rpt_family="FLAM_C"
repeat_region 23265. .23350
/rpt_family="(TA)n"
repeat_region 23357. .23381
/rpt_family="AT_rich"
repeat_region 24908. .25206
/rpt_family="AluSx"
repeat_region complement(25805. .25979)
/rpt_family="LIMB4A"
repeat_region complement(26043. .26262)
/rpt_family="LIMB4A"
repeat_region complement(26861. .26993)
/rpt_family="FLAM_C"
repeat_region complement(27023. .27317)
/rpt_family="AluSx"
repeat_region 27952. .28300
```

Query Match	Score	DB	Length
11.4%	113.8	9	167587

Best Local Similarity 55.1%; Pred. No. 3.4e-19;
 Matches 333; Constructed 0; Mismatches 100; Total 0; Count 0

Matches 223; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY	297	TGTCATATGGGGCCCAACCTGCAGAAATTAGGAAGCCCCAACCTCCCGAATACGAGGCCACC	356
Dd	119174	TGTTCACAAGCCTCTCTCTCTCTGTTCACGTAGCCTCTCTCTCTGTTCACAAGCCTCTCTC	119233
QY	357	TGCAGATATATGAGGCCCAACCCGTAGAAATAAGAGCCCGCCTGTGGGATPACAGAGCTTC	416
Dd	119234	TTCTGTTCACGTAGGCTCTCTCTCTGTTCACAGAGCTTCTCTCTGTTCACGTAGGCTTC	119293
QY	417	AAGTGCGAATATGAGGCCCAACCTCTTGGATACGAGGCCCAACTGCAGAAATATGAGC	476
Dd	119294	TCCTCTGTTCACAAGCCTCTCTCTGTTCACGTAGGCTCTCTCTGTTCACAAGC	119353
QY	477	CCCACCTCTAGGATATGAGGCCCAACCTTGGATATGAGAAGCCCACTCTGGAAATAGG	536
Dd	119354	CTCTCTCTGTTCATGTAGGCTCTCTCTCTGTTCATGTAGGCTCTCTCTGTTCACGT	119413
QY	537	AGCCCAACCTCTCGAATATGAGGCCCAACTGCAGAAATGAAAGGCCGCGTGGGGAATG	596
Dd	119414	AGCTCTCTCTCTGTTCACGTAGGCTCTCTCTGTTCACGTAGGCTCTCTCTGTTC	119473
QY	597	CAGAGCTTCACCTGTGATCAGAGGCCAGGCTTCAGAAATTCACAGAGCCAGGCTCC	656
Dd	119474	CACAGCTCTCTCTGTTCACGTAGGCTCTCTCTGTTCACAGAGCTCTCTCTGT	119533
QY	657	TGAAAACGAGGCTTCTCTCTCTGCTCTCTCTCTCTCATAGTCCA	701
Dd	119534	TCATGTGACCTCTCTCTCTGTTCATGTAGGCTCTCTCTCTGTTC	119578

Search completed: December 16, 2003, 15:10:18
Job time : 3730.91 secs

PD 29-NOV-2001.

PX 25-MAY-2001; 2001WO-CA00738.

XX

PR 25-MAY-2000; 2000CA-2307128.

PR 25-MAY-2000; 2000US-206979P.

PX

PA (TOOH) UNIT QUEENS KINGSTON.

PA (UYOR-) UNIT OREGON HEALTH SCI.

XX

Pi

Oko R, Sutovsky P;

XX

DR MPI: 2002-097644/13.

P-PsDB; AAU74610.

XX

PT Isolated perinuclear theca 32 polypeptide that interacts with activated

PT tyrosine kinase c-Yee, for enhancing fertility, treating/diagnosing

PT diminished fertility and abnormal spermiogenesis and for providing

PT contraception -

XX

PS Claim 62; Fig 4B; 103bp; English.

XX

CC The invention describes an isolated perinuclear theca 32 (PT32)

CC polypeptide (I) which interacts with tyrosine kinase c-Yee. (I) is

CC useful for: enhancing fertility in a mammal; creating globozoosperm, by

CC expressing (I) in spermatozoa; inhibiting fertilisation, by introducing

CC (I) or its antigenic fragment into a mammal to elicit an immune

CC response; enhancing the ability of round spermatids to activate oocytes;

CC treating or diagnosing diminished fertility and abnormal spermiogenesis;

CC in providing contraception; identifying contraceptive and

CC fertility-enhancing agents. The polynucleotide is useful for producing

CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and

CC for chromosome identification. An antibody against (I) is useful in

CC immunological assays, in immunocontraceptive methods, to identify cells

CC expressing (I), and to purify (I) by affinity chromatography. A

CC transgenic animal is useful as an animal model for studying human

CC fertility and reproductive biology, and for screening compounds to

CC identify modulators of oocyte activation. The use of (I) prevents the

CC entry of components which are detrimental to embryonic development into

CC the oocyte during oocyte activation with crude sperm extract and avoids

CC the propagation of viruses such as HIV (human immunodeficiency virus) and

CC SIV (Simian immunodeficiency virus) carried in the sperm. This sequence

CC encodes the human testicular WW domain binding protein (htMBP), described

CC in the method of the invention.

XX

SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other:

Query Match 100.0%; Score 1001; DB 24; Length 1001;
Best Local Similarity 100.0%; Pred. No. 5,3e-290;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Gy 1 ATGCCATTGATCTGTATGACGAACCTCATCTGTGGAAACAACAGTATTGCTCGAATTC 60
Db 1 ATGCCATTGATCTGTATGACGAACCTCATCTGTGGAAACAACAGTATTGCTCGAATTC 60

Gy 61 ATTAAAGGAACATTTACAGGACACTCCATATGAGTGCTGGGAAGAACAAGCTACTTTAAA 120
Db 61 ATTAAAGGAACATTTACAGGACACTCCATATGAGTGCTGGGAAGAACAAGCTACTTTAAA 120

Gy 121 TTAGTCTTCAGAAATGAGAGTGCATTGAAATTTGCCCATGTTGATGCTGCTCT 180
Db 121 TTAGTCTTCAGAAATGAGAGTGCATTGAAATTTGCCCATGTTGATGCTGCTCT 180

Gy 181 GCTGTGTCCTCAGAGATTTTCATCTTAGAACCTTAATATGATGATCTCTATGGAAT 240
Db 181 GCTGTGTCCTCAGAGATTTTCATCTTAGAACCTTAATATGATGATCTCTATGGAAT 240

Gy 241 TATGTATATTAATCTGGGGAAGGAATATATGACACTCCACAGATGCTTGATATATGTC 300
Db 241 TATGTATATTAATCTGGGGAAGGAATATATGACACTCCACAGATGCTTGATATATGTC 300

Gy 301 TATGGGCCCCCACCTGACAGATATGAGGCCCACTCCCGATACGAGCCCACCTGCA 360

Dd		301	TATGAGGGCCCGCACCTGTGACAGAAATATGAGACCACCACTCCCGGATACGAGAGGCCCACTGTGCA	368
Oy		361	GGAATATGAGAGCCCAACCCGTAGAGAAATAGAAGCCCGCCTGTGGGATACAGAGCTCACTT	420
Dd		361	GGATTATGAGAGCCCAACCCGTAGAGAAATAGAAGCCCGCCTGTGGGATACAGAGCTCACTT	420
Oy		421	GTGCAGATATGAGAGCCCAACTTTGTGAATACGAGGCCCACTTGACAGATATATGAGCCCCA	480
Dd		421	GTGCAGATATGAGAGCCCAACTTTGTGAATACGAGGCCCACTTGACAGATATATGAGCCCCA	480
Oy		481	CCTCTAGATATATGAGAGCCCAACTTGTGAATATGAGAACCCCACTCTCGATATATGAGACC	540
Dd		481	CCTCTAGATATATGAGAGCCCAACTTGTGAATATGAGAACCCCACTCTCGATATATGAGACC	540
Oy		541	CCAACCTTCGAGATATGAGAGCCCACTGTGACAGAAAATGAGAGCCCGCCTGTGGGATACAGA	600
Dd		541	CCAACCTTCGAGATATGAGAGCCCACTGTGACAGAAAATGAGAGCCCGCCTGTGGGATACAGA	600
Oy		601	GCCCTCACCTGTGTGAATACAGAGAGCCAGGCTCTAGGAAATTAACAGAGCCCAAGCTCCTGAA	660
Dd		601	GCCCTCACCTGTGTGAATACAGAGAGCCAGGCTCTAGGAAATTAACAGAGCCCAAGCTCCTGAA	660
Oy		661	AACGAGGCTCTCTTCCCCTGTGCGCCCTCTCTCTGACGGTCAATCTTAAACCTTCTAAGATG	720
Dd		661	AACGAGGCTCTCTTCCCCTGTGCGCCCTCTCTCTGACGGTCAATCTTAAACCTTCTAAGATG	720
Oy		721	TAAACCTTAGAGATCTACCAACAGCAAAGAGTAACTTAATAATTTGAAGTACAGATTAAGAGG	780
Dd		721	TAAACCTTAGAGATCTACCAACAGCAAAGAGTAACTTAATAATTTGAAGTACAGATTAAGAGG	780
Oy		781	ACGACTCAGCTTAGAGATCATTTGATGATCTGCAATTTGTGAAAAATTAGAGAAACAGATGCTC	840
Dd		781	ACGACTCAGCTTAGAGATCATTTGATGATCTGCAATTTGTGAAAAATTAGAGAAACAGATGCTC	840
Oy		841	CCATGTTCTCAAGAGAGACACTTTCTTAAGAGAAATCAATGCTGGGCAACAATATATGAGAG	900
Dd		841	CCATGTTCTCAAGAGAGACACTTTCTTAAGAGAAATCAATGCTGGGCAACAATATATGAGAG	900
Oy		901	GAAAGAGAGAGAGACACTCAACCAACTAGAGCAAGGATTAAGTTTTCCCTTGTTCACCTTTTC	960
Dd		901	GAAAGAGAGAGAGACACTCAACCAACTAGAGCAAGGATTAAGTTTTCCCTTGTTCACCTTTTC	960
Oy		961	AGTGTCTGCTGGAATGTATGATTAACCTCATTTGCAAGTAG	1001
Dd		961	AGTGTCTGCTGGAATGTATGATTAACCTCATTTGCAAGTAG	1001
RESULT 2				
ID	ABK84798		standard; cDNA; 220895 BP.	
XX	ABK84798;			
XX				
DT	14-AUG-2002	(first entry)		
XX				
DE			Human cDNA differentially expressed in granulocytic cells #1369.	
XX				
KM	Human; ss; granulocytic cell; DNA chip; bacterial infection;			
KM	viral infection; parasitic infection; protozoal infection;			
KM	fungal infection; sterile inflammation; diseases; poxiriasis;			
KM	rheumatoid arthritis; glomerulonephritis; asthma; thromboisis;			
KM	cardiac reperfusion injury; renal reperfusion injury; AIDS;			
KM	adult respiratory distress syndrome; inflammatory bowel disease;			
KM	Crohn's disease; ulcerative colitis; periodontal disease;			
KM	granulocyte activation; chronic inflammation; allergy.			
XX				
OS	Homo sapiens.			
XX				
PM	WO200228999-A2.			
XX				
PD	11-APR-2002.			
XX				
PF	03-OCT-2001; 2001WO-US30821.			

PR	14-AUG-2000	2000US-0228775
PR	18-AUG-2000	2000US-0228279
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226688
PR	23-AUG-2000	2000US-0221182
PR	30-AUG-2000	2000US-0229924
PR	01-SEP-2000	2000US-0228287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0228509
PR	05-SEP-2000	2000US-0228513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0230880
PR	08-SEP-2000	2000US-0230881
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233400
PR	14-SEP-2000	2000US-0233401
PR	14-SEP-2000	2000US-0233403
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	27-SEP-2000	2000US-0233484
PR	27-SEP-2000	2000US-0233584
PR	27-SEP-2000	2000US-0233586
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236870
PR	29-SEP-2000	2000US-0236872
PR	13-OCT-2000	2000US-0239355
PR	13-OCT-2000	2000US-0239357
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241185
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610

	PR	08-NOV-2000;	2000US-0246611.
	PR	08-NOV-2000;	2000US-0246613.
	PR	17-NOV-2000;	2000US-0249207.
	PR	17-NOV-2000;	2000US-0249208.
	PR	17-NOV-2000;	2000US-0249209.
	PR	17-NOV-2000;	2000US-0249210.
	PR	17-NOV-2000;	2000US-0249211.
	PR	17-NOV-2000;	2000US-0249212.
	PR	17-NOV-2000;	2000US-0249213.
	PR	17-NOV-2000;	2000US-0249214.
	PR	17-NOV-2000;	2000US-0249215.
	PR	17-NOV-2000;	2000US-0249216.
	PR	17-NOV-2000;	2000US-0249217.
	PR	17-NOV-2000;	2000US-0249218.
	PR	17-NOV-2000;	2000US-0249244.
	PR	17-NOV-2000;	2000US-0249245.
	PR	17-NOV-2000;	2000US-0249264.
	PR	17-NOV-2000;	2000US-0249265.
	PR	17-NOV-2000;	2000US-0249297.
	PR	17-NOV-2000;	2000US-0249299.
	PR	17-NOV-2000;	2000US-0249300.
	PR	01-DEC-2000;	2000US-0250160.
	PR	01-DEC-2000;	2000US-0250391.
	PR	05-DEC-2000;	2000US-0250130.
	PR	05-DEC-2000;	2000US-0251988.
	PR	05-DEC-2000;	2000US-0256719.
	PR	06-DEC-2000;	2000US-0251479.
	PR	08-DEC-2000;	2000US-0251856.
	PR	08-DEC-2000;	2000US-0251868.
	PR	08-DEC-2000;	2000US-0251869.
	PR	08-DEC-2000;	2000US-0251989.
	PR	08-DEC-2000;	2000US-0251990.
	PR	11-DEC-2000;	2000US-0254097.
	PR	05-JAN-2001;	2001US-0259678.
XX			
XX	(HUMA-)	HUMAN GENOME SCL INC.	
PA			
PI	Rosen CA,	Barash SC,	Ruben SM;
XX			
DR	WPI; 2001-	465570/50.	
P-PSDB;	AAM95259.		
XX			
PT	Isolated nucleic acid molecule encoding a reproductive system antigen		
PP	is used in preventing, treating or ameliorating a medical condition -		
XX			
PS	Claim 1;	SEQ ID NO 1230;	1297pp + Sequence Listing; English.
XX			
CC	The present invention provides the protein and coding sequences of a		
CC	number of human reproductive system related antigens. These can be used		
CC	in the prevention and treatment of reproductive system disorders,		
CC	including cancer. The present sequence is a coding sequence of the		
CC	invention.		
XX			
XX			
Seq	Sequence 467 BP;	109 A;	132 C;
			124 G;
			100 T;
			2 other;
	Query Match	44.6%;	Score 446.2;
	Best Local Similarity	98.5%;	Pred. No. 1.5e-123;
	Matches 459;	Conservative 1;	Mismatches 5;
			Indels 1;
			Gaps 1.
QY			
Db	129	CAGAAATGGAATGGCATTGCAATTTGCCAAGTGTATGATGAAGCTGCTGCTGTTC	188
	2	CAGAATGAGGTGCATTGAATTTGCCAGTTGATGTGAAGCTGCTGCTGCTGC	61
QY			
Db	189	CCGAGGATTTCCACTTAGAACCTTAATAAGACTGGTGAGCTCTAAGGAAATTATGAT	248
	62	CCGAGGATTTCCACTTAGAACCTTAATAAGACTGGTGAGCTCTAAGGAAATTATGAT	121
QY			
Db	249	TACTGGGGAAGGAATATGTGACATCCACAGATGCTGTTCAGTTATGTCTATGGGAC	308
	122	TACTGGGGAAGGAATATGTGACATCCACAGATGCTGTTCAGTTATGTCTATGGAGC	181
QY			
	309	CCCATCTGACAGATATGGAGCCCCACCTCCCGATACGAGACCACCTGCAGATATGG	368

Db 182 CCCACTGACAGATATGAGCCCACTCCCGATACGAGAGCCCACTGACGATATNG 241
QY 369 AGCCCAACCCCTAGAAATGAGAGCCCTGTGGATACAGAGCTCACTGTGGCTA 428
Db 242 AGCCCAACCCCTAGAAATGAGAGCCCTGTGGATACAGAGCTCACTGTGGCTA 301
QY 429 TGAAGCCCACTCTGTGATACGAGAGCCCACTGACGATATGAGAGCCCACTCTAG 488
Db 302 TGAAGCCCACTCTGTGATACGAGAGCCCACTGACGATATGAGAGCCCACTCTAG 361
QY 489 ATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTC 547
Db 362 ATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCT 421
QY 548 TCGATATGAGAGCCCACTGACGAGAAATGAGAGCCCACTGCGG 593
Db 422 TCGATATGAGAGCCCACTGACGAGAAATGAGAGCCCACTGCGG 467

RESULT 4
ABL96688
ID ABL96688 standard; cDNA; 467 BP.
XX
AC ABL96688;
XX
DT 21-JUN-2002 (first entry)
XX
DE Human testicular antigen encoding cDNA SEQ ID NO: 356.
XX
KM Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KM reproductive system disorder; urinary system disorder; gene therapy;
KM cardiovascular disorder; respiratory disorder; neurological disorder;
KM gastrointestinal disease; infection; cytostatic; gene; ss.
XX
OS Homo sapiens.
PN MO200155317-A2.
PD 02-AUG-2001.
XX
PP 17-JAN-2001; 2001WO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236357.
PR 29-SEP-2000; 2000US-0236358.
PR 29-SEP-2000; 2000US-0236359.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240660.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 06-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-483232/52.
 DR
 XX
 XX Nucleic acids encoding 973 human testicular antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating testicular cancer -
 XX
 PS Claim 1; SEQ ID NO 356; 766pp; English.
 XX
 XX The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a cDNA of the
 CC invention.
 CC
 XX Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
 SQ
 Query Match 44.6%; Score 446.2; DB 23; Length 467;
 Best Local Similarity 98.5%; Pred. No. 1.5e-123;
 Matches 459; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

QY 369 AGCCCAACCCGTAGAAATGAGGCCCGCTGTGGATACAGAGCCTCAGCTGTGGATA 428
 DB 242 AGCCCAACCCGTAGAAATGAGGCCCGCTGTGGATACAGAGCCTCAGCTGTGGATA 301
 QY 429 TGGAGCCCAACCTCTTGGATACGAGGCCCACTGCGAGATATGAGGCCCACTCTAG 488
 DB 302 TGGAGCCCAACCTCTTGGATACGAGGCCCACTGCGAGATATGAGGCCCACTCTAG 361
 QY 489 ATATGAGGCCCACTCTTGGATATGAGGCCCACTCTGAGATATGAGGCCCACTCTC 547
 DB 362 ATATGAGGCCCAACCTCTTGGATATGAGGCCCACTCTGAGATATGAGGCCCACTCT 421
 QY 548 TGGATATGAGGCCCACTGCGAGAAATGAGGCCCGCTGCGAG 593
 DB 422 TGGATATGAGGCCCACTGCGAGAAATGAGGCCCGCTTTCGAG 467
 RESULT 5
 AAS20601 standard; cDNA; 1413 BP.
 ID AAS20601;
 AC AAS20601;
 XX
 XX 09-APR-2002 (first entry)
 DT
 XX
 DE DNA encoding bovine perinuclear theca 32 (PT32).
 XX
 XX Testicular WW domain binding protein; hTWBP; perinuclear theca 32;
 KM PT32; contraceptive; fertility; oocyte activation; vaccine;
 KM globozoospermy; spermatogenesis; spermatozoa; tyrosine kinase; c-Yes;
 KM immunoc contraceptive; bovine; gene; ss.
 XX
 OS Bos sp.
 XX
 XX
 FH Key Location/Qualifiers
 XX primer_bind 30..50
 FT /tag= a
 FT /note= "Primer binding site for cDNA isolation. The
 FT sequence differs from that of the forward primer
 FT given in AAS20603"
 FT CDS 36..97
 FT /tag= b
 FT /product= "PT32"
 FT /note= "Perinuclear theca 32"
 FT protein_bind 978..1001
 FT /tag= c
 FT /note= "Primer binding site for cDNA isolation. The
 FT sequence differs from that of the reverse primer
 FT given in AAS20604"
 XX
 XX WO200190185-A2.
 XX
 XX 29-NOV-2001.
 XX
 XX 25-MAY-2001; 2001MO-CA00738.
 XX
 XX 25-MAY-2000; 2000CA-2307128.
 XX
 XX 25-MAY-2000; 2000US-206979P.
 XX
 XX (TOOH) UNIV QUEBENS KINGSTON.
 PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX
 XX
 PI Oko R, Sutovsky P;
 XX
 XX WPI; 2002-097644/13.
 DR P-PSDB; AAU74604.
 XX
 XX Isolated perinuclear theca 32 polypeptide that interacts with activated
 PT tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing
 PT diminished fertility and abnormal spermatogenesis and for providing
 PT contraception -

PS Claim 10; Fig 2A-B; 103bp; English.
 XX
 CC The invention describes an isolated perinuclear theca 32 (PT32)
 CC polypeptide (I) which interacts with tyrosine kinase c-188. (I) is
 CC useful for: enhancing fertility in a mammal; treating globoospermy, by
 CC expressing (I) in spermatozoa; inhibiting fertilization, by introducing
 CC (I) or its antigenic fragment into a mammal to elicit an immune
 CC response; enhancing the ability of round spermatids to activate oocytes;
 CC treating or diagnosing diminished fertility and abnormal spermiogenesis;
 CC in providing contraception; identifying contraceptive and
 CC fertility-enhancing agents. The polynucleotide is useful for producing
 CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
 CC for chromosome identification. An antibody against (I) is useful in
 CC immunological assays. In immun contraceptive methods, to identify cells
 CC expressing (I), and to purify (I) by affinity chromatography. A
 CC transgenic animal is useful as an animal model for studying human
 CC fertility and reproductive biology, and for screening compounds to
 CC identify modulators of oocyte activation. The use of (I) prevents the
 CC entry of components which are detrimental to embryonic development into
 CC the oocyte during oocyte activation with crude sperm extract and avoids
 CC the propagation of viruses such as HIV (human immunodeficiency virus) and
 CC SIV (simian immunodeficiency virus) carried in the sperm. This sequence
 CC encodes the bovine perinuclear theca 32 (PT32), described in the method
 CC of the invention.
 XX
 SQ Sequence 1413 BP; 377 A; 363 C; 369 G; 304 T; 0 other;
 Query Match 43.5%; Score 435.4; DB 24; Length 1413;
 Best Local Similarity 74.7%; Pred. No. 4.8e-120;
 Matches 609; Conservative 0; Mismatches 181; Indels 25; Gaps 4;
 QY 1 ATGCCATTGATCTGATGACCAACCTCACTGTGAAACCAAGTATTTGCTGCAACTTC 60
 DB 258 ATGCCGTTTGGCTGATGATGACCTGACCTGAAACCAATTTTGGCCCAATAC 317
 QY 61 ATTAAGGAACATATTCAGGAGCTCCATATGATGATGATGATGATGATGATGATGAT 120
 DB 318 ATTAAGGAACATATTCAGGAGCTCCATATGATGATGATGATGATGATGATGATGAT 377
 QY 121 TTAGCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
 DB 378 TTAGCTTCAGAAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 437
 QY 181 GCTGTGCGGAGATTTCCATTGAACTTTAAATGATGATGATGATGATGATGATGATGAT 240
 DB 438 GCTGTGCGGAGATTTCCATTGAACTTTAAATGATGATGATGATGATGATGATGATGAT 497
 QY 241 TATGTAATTAATGAGGAGAT---ATGCACTCCAGATGATGATGATGATGATGATGAT 292
 DB 498 TACATAATTAATGAGGAGAT---ATGCACTCCAGATGATGATGATGATGATGATGAT 557
 QY 293 ----TATATGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 348
 DB 558 CCAATGATGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 617
 QY 349 GCCCACCCTGAGGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 408
 DB 618 ACTCACAGGAGGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT 677
 QY 409 AGACCTCACTGAGGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 468
 DB 678 GAGACCTCACTGAGGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 737
 QY 469 TATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 528
 DB 738 TATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 797
 QY 529 GGAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 588
 DB 798 GGAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 857
 QY 589 GCGGATATCAGAGCTCACTGCTGATCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 648

DB 858 GCTGGAATGAGAGCCCTACCCCTGATATGAAAGCTCACTGCTGGAATACAGCTGCC 917
 QY 649 -----CAGGCTCTTGAAGAGAGGCTTCTTCCCTGCTGCTCTTCTGAC 696
 DB 918 TCTCAGAGATCTATGACAGCTCAGCAGAGAACTTCTTCCACTACTCATCTTCTTAG 977
 QY 697 GTCCATCTTAACCTTCTAAGATGTAAACCTTGAAGACTCAGCAAGAGAGGATACCT 756
 DB 978 GTCCATCTTAACCTTCTAAGATGTAAACCTTGAAGACTCAGCAAGAGAGGATACCT 1036
 QY 757 AAAATTGAGTCAAGATTAAGAGAGCAGCTCAGCT 791
 DB 1037 AAACTGAACTCAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 1071
 RESULT 6
 ID ABA67850 standard; DNA; 436 BP.
 XX
 AC ABA67850;
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #16155.
 XX
 KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 FN W020015727-A2.
 XX
 PD 09-AUG-2001.
 XX
 PP 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOL-B) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 4; SEQ ID NO 16155; 639bp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 43.4%; Score 434.4; DB 22; Length 436;
 Best Local Similarity 99.8%; Pred. No. 5.1e-120;
 Matches 435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 293 TTAGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 352
 DB 1 TTAGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60

QY 353 CACCTGCAAGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAG 412
 DB 61 CACCTGCAAGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAG 120
 QY 413 CCTGACCTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAG 472
 DB 121 CCTGACCTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAG 180
 QY 473 GAGCCCCACCTGTAGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 532
 DB 181 GAGCCCCACCTGTAGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 240
 QY 533 ATGAGAGCCCCACCTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 592
 DB 241 ATGAGAGCCCCACCTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 300
 QY 593 GATACAGAGCCCTCACTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 652
 DB 301 GATACAGAGCCCTCACTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 360
 QY 653 CTCCTGAAAACGAGGCTTCTCTGCTCTGCTCTGCTCTCTGAGTCAATCTTAACCTT 712
 DB 361 CTCCTGAAAACGAGGCTTCTCTGCTCTGCTCTGCTCTCTGAGTCAATCTTAACCTT 420
 QY 713 CTAAAGATGTAAACCTT 728
 DB 421 CTAAAGATGTAAACCTT 436

RESULT 7
 AAK42003
 ID AAK42003 standard; DNA; 436 BP.

XX AAK42003;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 16560.

KW Human; bone marrow expressed exon; gene expression analysis; probe;
 microarray; cancer; leukemia; lymphoma; myeloma; ss.

OS Homo sapiens.

XX WO200157276-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX Example 4; SEQ ID NO: 16560; 658bp + Sequence Listing; English.
 PS The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention.

Seq Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Query Match 43.4%; Score 434.4; DB 22; Length 436;
 Best Local Similarity 99.8%; Pred. No. 5.1e-120;
 Matches 435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 293 TTATTGTCTATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAGCC 352
 DB 1 TTATTGTCTATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAGCC 60
 QY 353 CACCTGCAAGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAG 412
 DB 61 CACCTGCAAGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAG 120
 QY 413 CCTGACCTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAG 472
 DB 121 CCTGACCTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGATACAGAG 180
 QY 473 GAGCCCCACCTGTAGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 532
 DB 181 GAGCCCCACCTGTAGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 240
 QY 533 ATGAGAGCCCCACCTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 592
 DB 241 ATGAGAGCCCCACCTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 300
 QY 593 GATACAGAGCCCTCACTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 652
 DB 301 GATACAGAGCCCTCACTGTGGATATGAGGCCCAACCTGTAGAAATGAAAGCCCCCTGTGGAT 360
 QY 653 CTCCTGAAAACGAGGCTTCTCTGCTCTGCTCTGCTCTCTGAGTCAATCTTAACCTT 712
 DB 361 CTCCTGAAAACGAGGCTTCTCTGCTCTGCTCTGCTCTCTGAGTCAATCTTAACCTT 420
 QY 713 CTAAAGATGTAAACCTT 728
 DB 421 CTAAAGATGTAAACCTT 436

RESULT 8

AAI48070
 ID AAI48070 standard; DNA; 436 BP.

XX AAI48070;

DT 17-OCT-2001 (first entry)

DE Probe #16756 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.

OS Homo sapiens.

XX WO200157272-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX DR WPI; 2001-488897/53.
 XX PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX PS Claim 25; SEQ ID No 16756; 654bp; English.
 CC CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 43.4%; Score 434.4; DB 22; Length 436;
 Best Local Similarity 99.8%; Pred. No. 5.1e-120;
 Matches 435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 293 TTATTTCTATGAGGCCCCCAGCTGAGATATGAGCCCACTCCGATACGAGGCC 352
 Db 1 TTATTTCTATGAGGCCCCCAGCTGAGATATGAGCCCACTCCGATACGAGGCC 60
 QY 353 CACCTGAGATATGAGGCCCAACCGTAGAATAAGAGCCCGCTGTGGATACAGAG 412
 Db 61 CACCTGAGATATGAGGCCCAACCGTAGAATAAGAGCCCGCTGTGGATACAGAG 120
 QY 413 CCTGACCTGATGATATGAGGCCCACTCTGATACGAGCCCACTGAGATATG 472
 Db 121 CCTGACCTGATGATATGAGGCCCACTCTGATACGAGCCCACTGAGATATG 180
 QY 473 GAGCCCACTCTGATATGAGGCCCACTCTGATATGAGGCCCACTCTGAGAT 532
 Db 181 GAGCCCACTCTGATATGAGGCCCACTCTGATATGAGGCCCACTCTGAGAT 240
 QY 533 ATGAGGCCCACTCTGATATGAGGCCCACTCTGAGATATGAGGCCCGCTGTGG 592
 Db 241 ATGAGGCCCACTCTGATATGAGGCCCACTCTGAGATATGAGGCCCGCTGTGG 300
 QY 593 GATACAGGCTCAGCTGATATGAGGCCCACTCTGATATGAGGCCCACTCTGAG 652
 Db 301 GATACAGGCTCAGCTGATATGAGGCCCACTCTGATATGAGGCCCACTCTGAG 360
 QY 653 CTCTGAAAACGAGAGCTTCTTCCCTCTGCTCTCTCTCTGAGATCTTAACTT 712
 Db 361 CTCTGAAAACGAGAGCTTCTTCCCTCTGCTCTCTCTCTGAGATCTTAACTT 420
 QY 713 CTAAGATGTAAACCTT 728
 Db 421 CTAAGATGTAAACCTT 436
 RESULT 9
 ABS16034
 ID ABS16034 standard; DNA; 436 BP.
 XX AC ABS16034;
 XX DT 19-AUG-2002 (first entry)
 XX DE Human genome-derived single exon probe ORF from lung SEQ ID No 16025.
 XX KM Human; de; single exon probe; asthma; lung cancer; COPD; ILD;
 KM chronic obstructive pulmonary disease; interstitial lung disease;
 KM familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KM tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KM Heremansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KM pulmonary histiocytosis; lymphangioleiomyomatosis; Kartagener syndrome;
 KM primary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KM primary ciliary dyskinesia; pulmonary hypertension;

KM hyaline membrane disease; open reading frame; ORF.
 XX OS Homo sapiens.
 XX XN WO200186003-A2.
 XX PD 15-NOV-2001.
 XX PP 30-JAN-2001; 2001WO-US00665.
 XX PR 04-FEB-2000; 2000US-180312P.
 XX PR 26-MAY-2000; 2000US-207456P.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-234687P.
 XX PR 27-SEP-2000; 2000US-236359P.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLB-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PT WPI; 2002-114183/15.
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX PS Claim 4; SEQ ID No 16025; 634bp; English.
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridization of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridization to a single exon
 CC microarray having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probe/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Heremansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Kartagener syndrome, fibrocystic
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_jct_sequences.
 XX SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 43.4%; Score 434.4; DB 24; Length 436;

Best Local Similarity 99.8%; Pred. No. 5,1e-120;
Matches 435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 293 TTATTGTCTATGAGGCCCCCACTGAGATATGAGAGCCCCCACTCCGGATACGAGAGCC 352
DB 1 TTATTGTCTATGAGAGCCCCCACTGAGATATGAGAGCCCCCACTCCGGATACGAGAGCC 60
QY 353 CACCTGAGATATGAGAGCCCCCACTGAGAAATGAGAGCCCCCGCTGAGATACAGAG 412
DB 61 CACCTGAGATATGAGAGCCCCCACTGAGAAATGAGAGCCCCCGCTGAGATACAGAG 120
QY 413 CCTGACCTGAGATATGAGAGCCCCCACTTGGATACGAGAGCCCCCACTGAGATATG 472
DB 121 CCTGACCTGAGATATGAGAGCCCCCACTTGGATACGAGAGCCCCCACTGAGATATG 180
QY 473 GAGCCCCCACTGAGATATGAGAGCCCCCACTTGGATATGAGAGCCCCCACTTGGAT 532
DB 181 GAGCCCCCACTGAGATATGAGAGCCCCCACTTGGATATGAGAGCCCCCACTTGGAT 240
QY 533 ATGAGAGCCCCCACTTGGATATGAGAGCCCCCACTTGGATATGAGAGAGCCCCCGCTGAG 592
DB 241 ATGAGAGCCCCCACTTGGATATGAGAGCCCCCACTTGGATATGAGAGAGCCCCCGCTGAG 300
QY 593 GATACAGAGCTGACCTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 652
DB 301 GATACAGAGCTGACCTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 653 CTCCTGAG 712
DB 361 CTCCTGAG 420
QY 713 CTAAAGTGTAACTT 728
DB 421 CTAAAGTGTAACTT 436

RESULT 10
AAL04882
ID AAL04882 standard; DNA; 7099 BP.
AC AAL04882;
XX
XX
DT 21-NOV-2001 (first entry)
XX
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 7570.
XX
XX
KM Human; reproductive system related antigen; reproductive system disorder;
XX
XX
OS Homo sapiens.
XX
XX
PN WO200155320-A2.
XX
XX
PD 02-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-019076.
PR 17-MAR-2000; 2000US-0198123.
PR 18-APR-2000; 2000US-0205515.
PR 19-MAY-2000; 2000US-0209467.
PR 07-JUN-2000; 2000US-0214886.
PR 28-JUN-2000; 2000US-0215135.
PR 30-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225271.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
XX
XX Disclosure; SEQ ID NO 7570; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a genomic sequence encoding a
XX protein of the invention.
XX
XX Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
Query Match 37.5%; Score 375.4; DB 22; Length 7099;
Best Local Similarity 99.7%; Pred. No. 1.2e-101;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 290 CAGTTATTGCTATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 349
|||||

Db 6723 CAGTTATTGCTATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 6782
QY 350 CCCACCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 409
Db 6783 CCCACCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 6842
QY 410 GAGCTCACTGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 469
Db 6843 GAGCTCACTGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 6902
QY 470 ATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 529
Db 6903 ATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 6962
QY 530 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 589
Db 6963 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 7022
QY 590 CGGATACAGAGCTCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 649
Db 7023 CGGATACAGAGCTCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAG 7082
QY 650 AGGCTCTGAAACGAG 666
Db 7083 AGGCTCTGAAACGAG 7099
RESULT 11
ABL97776
ID ABL97776 standard; DNA; 7099 BP.
XX
XX ABL97776;
DT 21-JUN-2002 (first entry)
XX
XX Human testicular antigen encoding DNA fragment SEQ ID NO: 2428.
DE
XX
XX Human, testicular antigen; testes; cancer; metastasis; immune disorder;
XX reproductive system disorder; urinary system disorder; gene therapy;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disease; infection; cytostatic; gene; ds.
OS Homo sapiens.
XX
XX MO20015317-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01329.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-MAR-2000; 2000US-0198123.
XX 19-MAR-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 14-JUL-2000; 2000US-0217496.
XX 26-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.

```

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225271.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.

```

```

PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251088.
PR 05-DEC-2000; 2000US-0255719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX MPI; 2001-483232/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer -
XX
XX PS Disclosure; SEQ ID NO 2428; 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a DNA encoding a
CC protein fragment of the invention.
XX
SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;

```

Query Match 37.5%; Score 375.4; DB 23; Length 7099;
Best Local Similarity 99.7%; Pred. No. 1,2e-101; Indels 0; Gaps 0;
Matches 376; Conservative 0; Mismatches 1;

```

QY 290 CAGTATGTCATGAGGCCCCACCTGACGATATGAGCCCCACCTCCCGATACGAG 349
DB 6723 CAGTATGTCATGAGGCCCCACCTGACGATATGAGCCCCACCTCCCGATACGAG 6782
QY 350 CCCACCTGACGATATGAGGCCCAACCGTATGAAATGAAAGCCCGCTGTGGATACA 409
DB 6783 CCCACCTGACGATATGAGGCCCAACCGTATGAAATGAAAGCCCGCTGTGGATACA 6842

```

QY 410 GAGCCTCACTGTGGATATGAGGCCCACTCTTGGATACGAGCCCCCACTGCGAGAT 469
DB 6843 GAGCCTCACTGTGGATATGAGGCCCACTCTTGGATACGAGCCCCCACTGCGAGAT 6902
QY 470 ATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGG 529
DB 6903 ATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGG 6962
QY 530 GATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGG 589
DB 6963 GATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGG 7022
QY 530 CGGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGG 649
DB 7023 CGGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTTGG 7082
QY 650 AGGCTCTGAAAACGAG 666
DB 7083 AGGCTCTGAAAACGAG 7099

RESULT 12
AA570582
ID AA570582 standard; cDNA; 894 BP.
AC AA570582;
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #6386.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
PN WO200175067-A2.
PD 11-OCT-2001.
PF 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ABG06395.
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID No 6386; 103pp; English.

The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridization probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABA54197-ABA54564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIR0
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 894 BP; 204 A; 221 C; 213 G; 256 T; 0 other;
Query Match 17.3%; Score 173.6; DB 23; Length 894;
Best Local Similarity 86.8%; Pred. No. 1.6e-41;
Matches 191; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 448 TACGAGCCCACTGCGAGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTT 507
DB 165 TATGAGGCCCACTGCGAGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTT 224
QY 508 GATATGAGGCCCACTGCGAGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTT 567
DB 225 GATATGAGGCCCACTGCGAGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTT 284
QY 568 GCAGAAATGAGGCCCACTGCGAGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTT 627
DB 285 GCAGAAATGAGGCCCACTGCGAGATATGAGGCCCACTCTTGGATATGAGGCCCACTCTT 344
QY 628 CCTCAGGAATCTACAGAGGCCCACTCTTGGAAACGAG 667
DB 345 CCTCAGGAATCTACAGAGGCCCACTCTTGGAAACGAG 384

RESULT 13
ABA55246
ID ABA55246 standard; DNA; 471 BP.
AC ABA55246;
DT 01-FEB-2002 (first entry)
DE Human foetal liver single exon nucleic acid probe #3551.
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
OS Homo sapiens.
PN WO200157277-A2.
PD 09-AUG-2001.
PF 30-JAN-2001; 2001WO-US00669.
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLR-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human foetal liver.
XX Claim 1; SEQ ID NO 3551; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WFO at ftp.wfo.int/pub/published_pct_sequences.
XX
SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;

Query Match 13.0%; Score 130.4; DB 22; Length 471;
Best Local Similarity 99.2%; Pred. No. 1.1e-28;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 290 CAGTTATGTCTATGAGGCGCCCACTGCAGATATGAGCCCACTCCCGATACGAG 349
DB 340 CAGTTATGTCTATGAGGCGCCCACTGCAGATATGAGCCCACTCCCGATACGAG 399
QY 350 CCCCACTGCAGATATGAGCCCACTGCAGATATGAGCCCACTCCCGATACGAG 409
DB 400 CCCCACTGCAGATATGAGCCCACTGCAGATATGAGCCCACTCCCGATACGAG 459
QY 410 GAGCCTCACCTG 421
DB 460 GAGCCTCACCTG 471

RESULT 14
AAK28957
ID AAK28957 standard; DNA; 471 BP.
XX
AC AAK28957;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 3514.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 3514; 658bp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.

XX
SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;

Query Match 13.0%; Score 130.4; DB 22; Length 471;
Best Local Similarity 99.2%; Pred. No. 1.1e-28;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 290 CAGTTATGTCTATGAGGCGCCCACTGCAGATATGAGCCCACTCCCGATACGAG 349
DB 340 CAGTTATGTCTATGAGGCGCCCACTGCAGATATGAGCCCACTCCCGATACGAG 399
QY 350 CCCCACTGCAGATATGAGCCCACTGCAGATATGAGCCCACTCCCGATACGAG 409
DB 400 CCCCACTGCAGATATGAGCCCACTGCAGATATGAGCCCACTCCCGATACGAG 459
QY 410 GAGCCTCACCTG 421
DB 460 GAGCCTCACCTG 471

RESULT 15
AAI34906
ID AAI34906 standard; DNA; 471 BP.
XX
AC AAI34906;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #3592 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488997/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 3592; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;

Query Match 13.0%; Score 130.4; DB 22; Length 471;
Best Local Similarity 99.2%; Pred. No. 1.1e-28;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 290 CAGTTATGTCTATGAGGCGCCCACTGCAGATATGAGCCCACTCCCGATACGAG 349

Db	340	CAATTATTGTCTATGAGCCCACTGAGATATGAGCCCACTCCCGGATTACGGAG	399
Qy	350	CCCCACCTGCGAGATATGAGCCCAACCCGTAGGAATGAAGCCCGCTGTGGATACA	409
Db	400	CCCCACCTGCGAGATATGAGCCCAACCCGTAGGAATGAAGCCCGCTGTGGATACA	459
Qy	410	GAGCCTGACCTG	421
Db	460	GAGCCTGACCTG	471

Search completed: December 16, 2003, 11:00:50
Job time : 270.133 secs

TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT. . .
FILE REFERENCE: XX/P04470US0
CURRENT APPLICATION NUMBER: US/08/841,349B
CURRENT FILING DATE: 1997-04-30
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 981
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
OTHER INFORMATION: For all n's in this sequence, n=(a or g or c or t)
US-08-841-349-15

Query Match 7.6%; Score 76; DB 2; Length 981;
Best Local Similarity 53.2%; Pred. No. 9.5e-14;
Matches 160; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 314 CTGCGAGATATGAGAGCCCACTCCCGGATACGAGAGCCCACTCGAGATATGAGGCC 373
DB 948 CTCGAGGCCATGTAGCCAGCTCTCCAGGCCATGTAGCCAGCTCTCCAGGCCATGTAGCCA 889
QY 374 AACCCGTAGGAATGAAGCCCGCTGTGGATACAGAGCTCACTGTGCGATATGAG 433
DB 888 GCTCTCAGGCCATGTAGCCAGCTCTCCAGGCCATGTAGCACTCTCCAGGTCATGTAG 829
QY 434 CCCACCTCTTGATACGAGAGCCCACTCGAGATATGAGAGCCCACTCTGATATG 493
DB 828 CCAAGCTCTCAGGCCATGTAGCCAGCTCTCAGGTCATGTAGCCAGCTCTCCAGGCCAT 769
QY 494 GACCCCACTCTTGATATGAGAGCCCACTCGGATATGAGAGCCCACTCTGAT 553
DB 768 TAGGAGATCTTCAGAGCATGTTGGCAGCTCTCCAGGATGTAGCCAGTTCTCCAGGCC 709
QY 554 ATGAGAGCCCACTCGAGGAATGAAGCCCGCTGTGGATACAGAGCTCACTGTG 613
DB 708 ATGTGGGAGCTCTCCAGGCCATGTAGCCAGCTCTCCAGGCCAATTTGGAGCTCTCCAG 649
QY 614 G 614
DB 648 G 648

RESULT 3
US-08-182-175A-56
Sequence 56, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Palco
APPLICANT: Sharon J. Keeler
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing R
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:

NAME: Linda Axamechy, Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DH5 alpha
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "gsp"
OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-182-175A-56

Query Match 5.4%; Score 54.4; DB 1; Length 243;
Best Local Similarity 53.2%; Pred. No. 2.4e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 302 ATGGGCCCCACCTGAGATATGAGAGCCCACTCCGGAATACGAGAGCCCACTGCA 361
DB 2 ATGAGAGAGAACTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAG 61
QY 362 GATATGAGAGCCCACTGAGGAATGAAGCCCGCTGTGGATACAGAGCTCACTG 421
DB 62 GCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTG 121
QY 422 TCGATATGAGAGCCCACTCTTGATACGAGAGCCCACTCGAGATATGAGAGCCCACT 481
DB 122 AAGCGATGAGAGAGAAAGCTTAAGAGATGAGAGAGAAAGCTTAAGATGAGAGAGAA 181
QY 482 CTCTAGATATGAGAGCCCACTCTTGATATGAGAA 517
DB 182 CTCAAAAGATGAGAGAAAGCTTAATGATGAGAA 217

RESULT 4
US-08-474-633A-74
Sequence 74, Application US/08474633A
Patent No. 5773691
GENERAL INFORMATION:
APPLICANT: COMPANY
APPLICANT: COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND
TITLE OF INVENTION: METHODS FOR INCREASING
TITLE OF INVENTION: INCREASING THE LYSINE
TITLE OF INVENTION: AND THREONINE CONTENT
TITLE OF INVENTION: OF THE SEEDS OF PLANTS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,633A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGEL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DH5 alpha
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic"
OTHER INFORMATION: storage protein
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "esp"
OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-474-633A-74

Query Match
Best Local Similarity 53.2%; Score 54.4; DB 1; Length 243;
Pred. No. 2.4e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 302 ATGGGGCCCCCAGTGAAGATGAGAGCCCACTCCCGATACGAGCCCACTGACG 361
|||
DB 2 ATGGAGGAGAAAGCTGAAGCGATGAGAGGAAAGCTGAAGCGATGAGAGGAAAG 61
|||
QY 362 GATATGAGAGCCCAAGCTGTAAGAAATGAAGCCCGCTGTGAGATACAGAGCCTG 421
|||
DB 62 GCGATGAGAGGAAAGCTGAAGCGATGAGAGGAAAGCTGAAGCGATGAGAGGAA 121
|||
QY 422 TCGATATGAGAGCCCACTCTTGGATACGAGAGCCCACTGAGATATGAGCCCA 481
|||
DB 122 AAGGCGATGAGAGGAAAGCTTAAGAAAGATGAGAGGAAAGCTGAATGATGAGAG 181
|||
QY 482 CTCTAGATATGAGAGCCCACTCTTGAATATGAA 517
|||
DB 182 CTCAAAAAGATGAGAGGAAAGCTTAATGATGAA 217
|||

RESULT 5
US-08-823-771-74
Sequence 74, Application US/08823771
Patent No. 6459019
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND
COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND
METHODS FOR INCREASING
INCREASING THE LYSINE
AND THREONINE CONTENT
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
AND COMPANY

STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/823,771
FILING DATE: 24-Mar-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/474,633
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGEL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DH5 alpha
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic"
OTHER INFORMATION: storage protein
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "esp"
OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-823-771-74

Query Match
Best Local Similarity 53.2%; Score 54.4; DB 4; Length 243;
Pred. No. 2.4e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 302 ATGGGGCCCCCAGTGAAGATGAGAGCCCACTCCCGATACGAGCCCACTGACG 361
|||
DB 2 ATGGAGGAGAAAGCTGAAGCGATGAGAGGAAAGCTGAAGCGATGAGAGGAAAG 61
|||
QY 362 GATATGAGAGCCCAAGCTGTAAGAAATGAAGCCCGCTGTGAGATACAGAGCCTG 421
|||
DB 62 GCGATGAGAGGAAAGCTGAAGCGATGAGAGGAAAGCTGAAGCGATGAGAGGAA 121
|||
QY 422 TCGATATGAGAGCCCACTCTTGGATACGAGAGCCCACTGAGATATGAGCCCA 481
|||
DB 122 AAGGCGATGAGAGGAAAGCTTAAGAAAGATGAGAGGAAAGCTGAATGATGAGAG 181
|||
QY 482 CTCTAGATATGAGAGCCCACTCTTGAATATGAA 517
|||
DB 182 CTCAAAAAGATGAGAGGAAAGCTTAATGATGAA 217
|||

RESULT 6
PCT-US92-06412-56
Sequence 56, Application PC/TUS9206412

GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing R
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E.I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06412
FILING DATE: 19920807
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamechy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "gsp"
OTHER INFORMATION: /standard_name= "7.7.7.7.8.9.8.9.5"
PCT-US92-06412-56
Query Match 5.4%; Score 54.4; DB 5; Length 243;
Best Local Similarity 53.2%; Pred. No. 2.4e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
Db 302 ATGGGGCCCCACCTGAGATATGAGCCCACTCCGAGATACGAGCCCACTGACG 361
2 ATGAGGAGAAAGCTTAAGGCGATGAGAGAAAGCTGAAGGCGATGAGAGAAAGCTGAAG 61
362 GATATGAGCCCAACCCGATGAGAAATGAAGCCCGCTGTGATACAGAGCTCACTG 421
62 GCGATGAGAGAAAGCTGAAGGCGATGAGAGAAAGCTGAAGGCGATGAGAGAAAGCTG 121
422 TGCATATGAGAGCCCACTCTTGTGATACGAGCCCACTGCAAGATATGAGAGCCCAAC 481
122 AAGGAGATGAGAGAAAGCTTAAGAAAGTGAAGAAAGCTGAATGATGAGAGAGAAA 181
482 CTCTAGATATGAGAGCCCACTCTTGTGATATGAA 517
182 CTCAAAAGATGAGAGAAAGCTTAATGATGAA 217

RESULT 7
US-08-917-320-18
Sequence 18, Application US/08917320
Patent No. 5824508
GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: No. 5824508 Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/917,320
FILING DATE: 25-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Luann Coert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
TELEX: 380816 COOLEYPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
US-08-917-320-18
Query Match 5.4%; Score 53.6; DB 1; Length 3833;
Best Local Similarity 50.4%; Pred. No. 2.3e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
Db 307 GCCCACCCTGCGAGATATGAGCCCACTCCGAGATACGAGCCCACTGCGAGATAT 366
2578 GCCCACCCTGCGAGATATGAGCCCACTCCGAGATACGAGCCCACTGCGAGATAT 2637
367 GAGCCCAACCCGATGAGAAATGAAGCCCGCTGTGATACAGAGCTCACTGTGGA 426
2638 CTACCCCAACCCCAATGAGCCCACTGCGAGAAACAAAGTCTTACCTACAGAG 2697
427 TATGAGCCCACTCTTGTGATATGAGCCCACTGCGAGATATGAGCCCACTGTA 486
2698 TGACTACCCCAACCCCAATGAGCCCACTGCGAGATATGAGCCCACTGTA 2757
487 GATATGAGCCCACTCTTGTGATATGAGCCCACTGCGATATGAGCCCACT 546
2758 CAGTGACTACCCCAACCCCAATGAGCCCACTGCGATATGAGCCCACT 2817
547 CTGGAATATGAGCCCACT 566
2818 CAGCAGTGACTACCCCACT 2837

```
RESULT 8
PCT-US95-04611A-18
; Sequence 18, Application PC/TUS9504611A
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
; TITLE OF INVENTION: Non Splicing Variants of gp350/220
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: 5 Palo Alto Square
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04611A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,291
; FILING DATE: April 18, 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cseert
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: AVIR-003/00US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-843-5163
; TELEFAX: 415-857-0663
; TELEX: 380816 CooleyPA
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3833 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1014..3734
; OTHER INFORMATION:
; PCT-US95-04611A-18

Query Match          5.4%; Score 53.6; DB 5; Length 3833;
Best Local Similarity 50.4%; Pred. No. 2.3e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 307 GCCCACTGACGATATGAGCCCACTCCCGATACGAGCCCACTGACAGATAT 366
DB 2578 GCCCACTGACGATATGAGCCCACTCCCGATACGAGCCCACTGACAGATAT 2637
QY 367 GAGGCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 426
DB 2638 CTACCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 2697
QY 427 TATGAGCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 486
DB 2698 TACTACCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 2757
QY 487 GATATGAGCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 546
DB 2758 CAGTACCTACCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 2817
QY 547 CTCGATATGAGAGCCCACT 566
DB 2818 CAGCAGTACCTACCCCACT 2837
```

```
RESULT 9
US-08-783-774-1
; Sequence 1, Application US/08783774
; Patent No. 6054130
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; TITLE OF INVENTION: Jackman, Winthrop
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/783,774
; FILING DATE: 15-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Cortuzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7682-037
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5931 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1014..3734
; OTHER INFORMATION:
; US-08-783-774-1

Query Match          5.4%; Score 53.6; DB 3; Length 5931;
Best Local Similarity 50.4%; Pred. No. 3e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 307 GCCCACTGACGATATGAGCCCACTCCCGATACGAGCCCACTGACAGATAT 366
DB 2578 GCCCACTGACGATATGAGCCCACTCCCGATACGAGCCCACTGACAGATAT 2637
QY 367 GAGGCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 426
DB 2638 CTACCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 2697
QY 427 TATGAGCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 486
DB 2698 TACTACCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 2757
QY 487 GATATGAGCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 546
DB 2758 CAGTACCTACCCCACTGATGAGAAATGAGAGCCGCTGTGGATACAGAGCTCACTGTGCA 2817
QY 547 CTCGATATGAGAGCCCACT 566
DB 2818 CAGCAGTACCTACCCCACT 2837
```

RESULT 10

US-09-556-706B-1
; Sequence 1, Application US/09556706B
; Patent No. 6458364
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Jackman, Winthrop
; TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220
; FILE REFERENCE: 7682-050-999
; CURRENT APPLICATION NUMBER: US/09/556,706B
; PRIOR FILING DATE: 2000-04-24
; PRIOR APPLICATION NUMBER: 08/783,774
; PRIOR FILING DATE: 1997-01-15
; PRIOR APPLICATION NUMBER: 08/229,291
; PRIOR FILING DATE: 1994-04-18
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 5931
; TYPE: DNA
; ORGANISM: Virus
; FEATURE:
; OTHER INFORMATION: GP350/220
US-09-556-706B-1

Query Match 5.4%; Score 53.6; DB 4; Length 5931;
Best Local Similarity 50.4%; Pred. No. 3e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 307 GCGCCACCTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 366
DB 2578 GCGCCACCTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 2637
QY 367 GGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 426
DB 2638 GCGCCACCTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 2637
QY 427 TATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 486
DB 2638 TATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 2757
QY 487 GATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 546
DB 2758 GATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 2817
QY 547 CTCGATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 566
DB 2818 CTCGATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 2837

RESULT 11
US-07-865-662F-10
; Sequence 10, Application US/07865662F
; Patent No. 5451670
; GENERAL INFORMATION:
; APPLICANT: Marcia M. Miller
; TITLE OF INVENTION: Restriction Fragment Length
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSER: City of Hope
; STREET: 1500 East Duarte Road
; CITY: Duarte
; STATE: California
; COUNTRY: United States of America
; ZIP: 91010-0269
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3M Double Density 5 1/4" diskette
; OPERATING SYSTEM: MS DOS Version 3.20
; SOFTWARE: Microsoft
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/865,662F
; FILING DATE: 07 April, 1992

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/688,326
FILING DATE: 22 April 1991
APPLICATION NUMBER: 07/588,922
FILING DATE: 27 September 1990
APPLICATION NUMBER: 07/210,405
FILING DATE: 23 June 1988
APPLICATION NUMBER: US 07/130,529
FILING DATE: 9 December 1987
APPLICATION NUMBER: US 07/068,176
FILING DATE: 30 June 1987
ATTORNEY/AGENT INFORMATION:
NAME: Irons, Edward S.
REGISTRATION NUMBER: 16,541
REFERENCE/DOCKET NUMBER: NO. 5451670e
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 783-6040
TELEFAX: (202) 783-6031
TELEX: No. 5451670e
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE: Synthetically Prepared
IMMEDIATE SOURCE: Synthetically Prepared
US-07-865-662F-10

Query Match 5.2%; Score 52; DB 1; Length 2188;
Best Local Similarity 47.5%; Pred. No. 5.1e-06;
Matches 154; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 302 ATGGGCCCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 361
DB 1315 ATGGGCCCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 1374
QY 362 GATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 421
DB 1375 GATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 1434
QY 422 TGCGATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 481
DB 1435 TGCGATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 1494
QY 482 CTCTAGATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 541
DB 1495 CTCTAGATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 1554
QY 542 CACTCTGAGATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 601
DB 1555 CACTCTGAGATATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 1614
QY 602 CCGACCTGCTGATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 625
DB 1615 CCGACCTGCTGATGAGCCCACTGAGATATGAGCCCACTCCCGATACGAGCCCACTGCAAGATAT 1638

RESULT 12
US-08-374-219B-10
; Sequence 10, Application US/08374219B
; Patent No. 6218106
; GENERAL INFORMATION:
; APPLICANT: Miller, Marcia M.
; TITLE OF INVENTION: Restriction Fragment Length Polymorphism Test
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Rothwell, Pigg, Ernst & Kurz
; STREET: 555 Thirteenth Street, N.W.
; CITY: Washington

```
STATE: D. C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/06/374,219B
FILING DATE: 18-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/865,662
FILING DATE: 07-APR-1992
APPLICATION NUMBER: 07/688,326
FILING DATE: 22-APR-1991
APPLICATION NUMBER: 07/588,922
FILING DATE: 27-SEP-1990
APPLICATION NUMBER: 07/413,301
FILING DATE: 28-SEP-1989
APPLICATION NUMBER: 07/210,405
FILING DATE: 23-JUN-1988
APPLICATION NUMBER: 07/130,529
FILING DATE: 09-DEC-1987
APPLICATION NUMBER: 07/068,176
FILING DATE: 30-JUN-1987
ATTORNEY/AGENT INFORMATION:
NAME: Walker, Barbara W
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 2124-113A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 783-6040
TELEFAX: 202 783-6031
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
US-08-374-219B-10

Query Match
Best Local Similarity 5.2%; Score 52; DB 3; Length 2188;
Best Local Similarity 47.5%; Pred. No. 5.1e-06;
Matches 154; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

302 ATGGGGCCCCCAGATATGAGCCCACTCCGGATTCGAGCCCCCACTGCAG 361
1315 ATGGGGCAACAGCTTAAGATCAGAGCAAGAAATCGAGCTGAAGAGCACATGAG 1374
362 GATATGAGCCCAACCGTAGAATGAAGCCCGCTGTGGATACAGAGCTCACTG 421
1375 GAGATGGGCAACAGCTTAAGATCAGAGCAAGAAATCGAGCTGAAGAGCACAT 1434
422 TCGATATGAGCCCACTCTTGATACGAGAGCCCACTGCAGATATGAGCCCA 481
1435 GAGGATGGGCAACAGCTTAAGATCAGAGCAAGAAATCGAGCTGAAGAGCAC 1494
482 CTCTAGATATGAGCCCACTCTTGATATGAGAAACCACTCTCGATATGAGCC 541
1495 CATGAGGATGGGCAACAGCTTAAGATCAGAGCAAGAAATCGAGCTGAAGAG 1554
542 CACTCTCGATATGAGCCCACTGCAGAGAAATGAAGCCCGCTGCGGATACAG 601
1555 CACCATGAGAGATGGGCAACAGCTTAAGATCAGAGCAAGAAATCGAGCTGA 1614
602 CCTCAGCTCTGATACAGAGCA 625
1615 GAGCACCATGAGAGATGGGCA 1638

RESULT 13
```

```
US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MS/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
FEATURES:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match
Best Local Similarity 5.2%; Score 52; DB 2; Length 3489;
Best Local Similarity 48.1%; Pred. No. 6.8e-06;
Matches 148; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

316 GAGGATATGAGCCCACTCCGGATTCGAGCCCACTGCAGATATGAGCCCA 375
2202 GAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAG 2261
376 CCGTAGAATGAAGCCCGCTGTGGATACAGAGCTCACTGTGATATGAGCC 435
2262 GAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGAGATTAGAG 2321
436 CCACTCTTGATATGAGAGCCCACTGCAGATATGAGAGCCCACTCTAGAT 495
2322 GAGATTAGAGATCAGAGCAGAGATTAGAGCAGCAGAGATTAGAGCAG 2381
496 GCCCACTCTTGATATGAGAACCCCACTCTCGATATGAGCCCACTCTGAT 555
2382 GAGAGATTAGAGAGCAGAGCAGAGATTAGAGAGCAGAGATTAGAGAG 2441
556 GAGGCCCACTGCAGAGAAATGAAGCCCGCTGCGGATACAGAGCTCACT 615
2442 GAGCAGAGATTAGAGAGCAGAGCAGAGATTAGAGAGCAGAGATTAGAG 2501
616 TCAGAGC 623
```



```

QY 309 CCCACTGAGATATGAGAGCCCACTCCGGATAGAGAGCCCACTGAGAGATAG 368
DB 182 CCCACTGAGATATGAGAGCCCACTCCGGATAGAGAGCCCACTGAGAGATAG 241
QY 369 AGCCCAACCCGATGAGAAATGAGAGCCCGCTGTGGATACAGAGCTCACTGTGGATA 428
DB 242 AGCCCAACCCGATGAGAAATGAGAGCCCGCTGTGGATACAGAGCTCACTGTGGATA 301
QY 429 TGGAGCCCACTCTTGGATACAGAGCCCACTGAGAGATATGAGAGCCCACTCTTAG 488
DB 302 TGGAGCCCACTCTTGGATACAGAGCCCACTGAGAGATATGAGAGCCCACTCTTAG 361
QY 489 AATGAGAGCCCACTCTTGGATACAGAGCCCACTGAGAGATATGAGAGCCCACTCT 547
DB 362 AATGAGAGCCCACTCTTGGATACAGAGCCCACTGAGAGATATGAGAGCCCACTCT 421
QY 548 TGGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGGAG 593
DB 422 TGGATATGAGAGCCCACTGAGAGATATGAGAGCCCGCTGTGGAG 467

```

RESULT 2

```

US-09-864-761-31561
; Sequence 31561, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117

```

```

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31561
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 299716.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
; OTHER INFORMATION: EST HUMAN HIT: AM169980.1, EVALUE 8.00e-11
; OTHER INFORMATION: SWISSPROT HIT: P18616, EVALUE 3.00e-13
US-09-864-761-31561

```

```

Query Match 43.4%; Score 434.4; DB 9; Length 436;
Best Local Similarity 99.8%; Pred. No. 2.1e-133;
Matches 435; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 293 TTAATGTCTATGAGGCGCCCACTGAGATATGAGAGCCCACTCCGGATACGAGGCC 352
DB 1 TTAATGTCTATGAGGCGCCCACTGAGATATGAGAGCCCACTCCGGATACGAGGCC 60
QY 353 CACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGTGGATACAGAG 412
DB 61 CACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGTGGATACAGAG 120
QY 413 CTTCACTGTGTGATATGAGAGCCCACTTGTGATATGAGAGCCCACTGTGATATG 472
DB 121 CTTCACTGTGTGATATGAGAGCCCACTTGTGATATGAGAGCCCACTGTGATATG 180
QY 473 GAGCCCACTCTAGATATGAGAGCCCACTTGTGATATGAGAGCCCACTCTGTGAT 532
DB 181 GAGCCCACTCTAGATATGAGAGCCCACTTGTGATATGAGAGCCCACTCTGTGAT 240
QY 533 ATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGTGG 592
DB 241 ATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCGCTGTGG 300
QY 593 GATACAGAGCTCACTGTGATACAGAGCCCGCTGAGATATGAGAGCCCGCTGTGG 652
DB 301 GATACAGAGCTCACTGTGATACAGAGCCCGCTGAGATATGAGAGCCCGCTGTGG 360
QY 653 CTCCTGAAAGAGGCTTCTCCCTGCTGCTCTCTGAGATATGAGAGCCCACTCTTGA 712
DB 361 CTCCTGAAAGAGGCTTCTCCCTGCTGCTCTCTGAGATATGAGAGCCCACTCTTGA 420
QY 713 CTAAGATGTAACCTT 728
DB 421 CTAAGATGTAACCTT 436

```

RESULT 3

```

US-09-764-891-7570
; Sequence 7570, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; Prior application data removed - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7570
; LENGTH: 7099
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7570

```

```

Query Match 37.5%; Score 375.4; DB 11; Length 7099;
Best Local Similarity 99.7%; Pred. No. 5.1e-113;

```

Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 290 CAGTTATTGCTATAGGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 349
DB 6723 CAGTTATTGCTATAGGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 6782

QY 350 CCCCACCTGAGATATGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 409
DB 6783 CCCCACCTGAGATATGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 6842

QY 410 GAGCCTCACTGAGATATGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 469
DB 6843 GAGCCTCACTGAGATATGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 6902

QY 470 ATGAGACCCCACTCTAGAGATATGAGCCCACTCTGAGATATGAGCCCACTCTGAG 529
DB 6903 ATGAGACCCCACTCTAGAGATATGAGCCCACTCTGAGATATGAGCCCACTCTGAG 6962

QY 530 GATATGAGCCCACTCTGAGATATGAGCCCACTCTGAGATATGAGCCCACTCTGAG 589
DB 6963 GATATGAGCCCACTCTGAGATATGAGCCCACTCTGAGATATGAGCCCACTCTGAG 7022

QY 590 CGGATACAGAGCCCTCACTGAGATATGAGCCCACTCTGAGATATGAGCCCACTCTGAG 649
DB 7023 CGGATACAGAGCCCTCACTGAGATATGAGCCCACTCTGAGATATGAGCCCACTCTGAG 7082

QY 650 AGGCTCTGAAAAAGAG 666
DB 7083 AGGCTCTGAAAAAGAG 7099

RESULT 4
US-09-864-761-15029

Sequence 15029, Application US/09864761

Patent No. US20020648763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wenheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecmika-X-1

CURRENT FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 09/864,761

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Anomax Sequence Listing Engine vers. 1.1

SEQ ID NO 15029

LENGTH: 471

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

OTHER INFORMATION: MAP TO 299716.4

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2

US-09-864-761-15029

Query Match 13.0%; Score 130.4; DB 9; Length 471;
Best Local Similarity 99.2%; Pred. No. 2.5e-32;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 290 CAGTTATTGCTATAGGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 349
DB 340 CAGTTATTGCTATAGGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 399

QY 350 CCCCACCTGAGATATGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 409
DB 400 CCCCACCTGAGATATGAGCCCACTGAGAGATATGAGCCCACTCCGAGATACGAG 459

QY 410 GAGCCTCACTGAG 421
DB 460 GAGCCTCACTGAG 471

RESULT 5

US-09-764-891-1356

Sequence 1356, Application US/09764891

Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

PRIOR FILING DATE: 2001-01-17

Prior application data removed - consult PAM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1356

LENGTH: 321

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: SITE

LOCATION: (316)

OTHER INFORMATION: n equals a,t,g, or c

NAME/KEY: SITE

LOCATION: (321)

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

OTHER INFORMATION: n equals a,t,g, or c

Qy 61 ATTAAAGGAAGTATTGAGGAGCTGATATGATGCTGGG-AGAGCAAGTACTTTAA 119
 Db 255 ATTAAAGGAAGTATTGAGGAGCTGATATGATGCTGGGAGAGCAAGTACTTTAA 314
 Qy 120 ATTA 123
 Db 315 AATA 318

RESULT 6
 US-09-764-891-7572
 ; Sequence 7572, Application US/09764891
 ; Publication No. US20030077808A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 7572
 ; LENGTH: 5769
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-764-891-7572

Query Match 11.0%; Score 110.2; DB 11; Length 5769;
 Best Local Similarity 90.1%; Pred. No. 6.5e-25;
 Matches 118; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 Qy 860 CTTCTTAAGGAAGTCACTGAGGCAAGTATGAGAGGAGAGAGACTCAA 919
 Db 3540 CTTACTTCCAACTACAGTACGTGGCAACAGTATGAGAGAGAGAGACTCAA 3599
 Qy 920 CCAACTAG 979
 Db 3600 CCAACTAG 3659
 Qy 980 TGATTAAGTCA 990
 Db 3660 TGATTAAGTCA 3670

RESULT 7
 US-10-029-386-22839/c
 ; Sequence 22839, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: ABOMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 22839
 ; LENGTH: 866
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC005630.1
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
 ; OTHER INFORMATION: SWISSPROT HIT: Q08379, EVALU 4.00e-10
 ; OTHER INFORMATION: NT HIT: G16161055, EVALU 4.00e-95
 ; OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALU 0.00e+00
 US-10-029-386-22839

Query Match 10.9%; Score 109; DB 13; Length 866;
 Best Local Similarity 54.3%; Pred. No. 4.9e-25;
 Matches 220; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

Qy 297 TGTCTATGAGGAGCCCACTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAG 356
 Db 577 TGTTCACAGAGCT 518
 Qy 357 TGCAGATATGAGAGCCCACTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAG 416
 Db 517 TGTCTATGAG 458
 Qy 417 ACCGTGAGATATGAG 476
 Db 457 TGTCTATGAG 398
 Qy 477 CCCACTTATGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 536
 Db 397 CT 338
 Qy 537 AGCCCACT 596
 Db 337 AGCT 278
 Qy 597 CAGAGCT 656
 Db 277 CAGAGCT 218
 Qy 657 TGAAG 701
 Db 217 TCAATGAGCT 173

RESULT 8
 US-10-029-386-26033/c
 ; Sequence 26033, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: ABOMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029,386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 26033
 ; LENGTH: 1040
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO AC005630.1
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
 ; OTHER INFORMATION: NT HIT: G16161055, EVALU 2.00e-97
 ; OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALU 0.00e+00
 ; OTHER INFORMATION: SWISSPROT HIT: Q08379, EVALU 6.00e-10
 US-10-029-386-26033
 Query Match 10.0%; Score 100; DB 13; Length 1040;
 Best Local Similarity 53.3%; Pred. No. 5.4e-22;
 Matches 211; Conservative 0; Mismatches 185; Indels 0; Gaps 0;
 Qy 306 GAGCCCACTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365
 Db 568 GAGAGCT 509
 Qy 366 TGAAG 425
 Db 508 CAGAGCT 449

```

QY 426 ATATGAGCCCACTCTTGATAGAGAGCCCACTGAGATATGAGCCCACTCT 485
    |||||
DB 448 TCAGTAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 389
    |||||
QY 486 AGATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTCTCTCTCTCTCT 545
    |||||
DB 388 TGTTCAACAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 329
    |||||
QY 546 TCTTGATATGAGAGCCCACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 605
    |||||
DB 328 TCTGTTCAAGTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 269
    |||||
QY 606 ACCTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 665
    |||||
DB 268 TCTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 209
    |||||
QY 666 GAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 701
    |||||
DB 208 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 173
    |||||

```

RESULT 9

```

US-09-764-891-7571
; Sequence 7571, Application US/09764891
; Publication No. US2003007808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7571
; LENGTH: 291
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7571

```

```

Query Match          9.3%; Score 93.2; DB 11; Length 291;
Best Local Similarity 96.9%; Pred. No. 4.3e-20;
Matches 95; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 ATGCCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
    |||||
DB 61 ATGCCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
    |||||
QY 61 ATTAAAGGAATATTTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 98
    |||||
DB 121 ATTAAAGGAATATTTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 158
    |||||

```

RESULT 10

```

US-09-864-761-4562/C
; Sequence 4562, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David R.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03

```

```

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4562
; LENGTH: 1984
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL009178.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
US-09-864-761-4562

```

```

Query Match          8.3%; Score 89.4; DB 9; Length 1984;
Best Local Similarity 51.6%; Pred. No. 2.7e-18;
Matches 204; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

```

```

QY 300 CTATGGAGCCCACTCTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
    |||||
DB 872 CTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 813
    |||||
QY 360 AGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
    |||||
DB 812 CAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 753
    |||||
QY 420 TGTGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479
    |||||
DB 752 TCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 693
    |||||
QY 480 ACCTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 539
    |||||
DB 692 ACCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 633
    |||||
QY 540 CCAGCCTCTGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 599
    |||||
DB 632 GACTCTCTCAAGTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 573
    |||||

```

QY 600 AGCTCACTGCTGATGAGAGCCAGCTCAGGAATGTAGAGAGCCAGCTCTCGA 659
DB 572 AAGAGCTCTCCCGAGTTCGGAGAGCTCTCCAGATTCTGAGAGACTCTCTCATTC 513
QY 660 AAACGAGCT 694
DB 512 CAGAGAGACTCTCTCCAGCTCTCCGAGAGAGACTCTCTC 478

RESULT 11
US-10-029-386-12333/c

Sequence 12333, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: A60MICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 12333
LENGTH: 541
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005630.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
OTHER INFORMATION: NT HIT: g413651705, EVALU8 2.00e-06
OTHER INFORMATION: EST_HUMAN HIT: BF000754.1, EVALU8 0.00e+00
US-10-029-386-12333

Query Match Best Local Similarity 8.9%; Score 88.8; DB 13; Length 541;

Matches 111; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 306 GGGCCCACTGAGATATGAGAGCCCACTCCGAGATAGAGAGCCCACTGAGAGAT 365
DB 314 GAGAGCT 255
QY 366 TGAAGCCCAAGCCGTAAGAAATGAGAGCCCGCTGAGAGATACAGAGCTCTCTG 425
DB 254 CAGAGCT 195
QY 426 ATATGAGAGCCCACTCTTGTGATACGAGAGCCCACTGAGAGATATGAGAGCCCACTCT 485
DB 194 TCAAGTACCT 135
QY 486 AGGATATGAGAGCCCACTCTTGTGATATGAGAGCCCACTCTGAGATATGAGAGCCCACT 545
DB 134 TGTTCACAGAGCT 75
QY 546 TCTGAGATATGAGAGCCCACTGAGAGAAATGAGAGCCCGCTGAGAGATACAGAGCTCT 605
DB 74 TCTGTTTACAGTACCT 15
QY 606 ACCTGCTG 613
DB 14 TCTCTCATG 7

RESULT 12

US-10-029-386-24981
Sequence 24981, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.

FILE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: A60MICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 24981
LENGTH: 511
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO 297205.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
US-10-029-386-24981

Query Match Best Local Similarity 8.8%; Score 87.8; DB 13; Length 511;

Matches 194; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 283 CTTGTTGATATGTTCTATAGAGGAGCCCACTGAGAGATATGAGAGCCCACTCCGGA 342
DB 138 CTTCTTCACTGAGAGCT 197
QY 343 TACGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGAAATGAGAGCCCGCTGTG 402
DB 198 TCCCTCTTCACTGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 257
QY 403 GATATGAGAGCT 462
DB 258 GAGTCT 317
QY 463 GAGAGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCCCACT 522
DB 318 GAGAGCT 377
QY 523 CTTCTGATATGAGAGCCCACTCTCTGATATGAGAGCCCACTGAGAGAAATGAGAGC 582
DB 378 CTTGAGAGCT 437
QY 583 CCGCTGAGAGATACAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 642
DB 438 CCACTGAGAGCT 497
QY 643 GAGGCCAGGC 653
DB 498 CTTCACTATGC 508

RESULT 13

US-09-864-761-21308/c
Sequence 21308, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: A60MICA-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21308
LENGTH: 543
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL009178.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
OTHER INFORMATION: EST_HUMAN HIT: AM382320.1, EVALUE 8.80e-02
OTHER INFORMATION: SWISSPROT HIT: Q05904, EVALUE 7.00e-12
US-09-864-761-21308

Query Match 8.7%; Score 87; DB 9; Length 543;

Best Local Similarity 54.1%; Pred.No. 7.4e-18; Matches 177; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 305 GGGCCCCCCTGAGATATGAGCCCACTCCGAGATGAGAGCCCACTGAGAT 364
DB 403 GGGAACTCTCTCCAGTTCCGAGAGACCTTCCAGCTCCGAGAGACTCTCCAGCT 344
QY 365 ATGAGCCCAACCCGTAGAAATGAAGCCCGCTGTGGATACAGAGCTCACTGTGC 424
DB 343 CCGAGAGAGACTCTCCAGTTCCGAGAGACTCTCCAGCTCCAGAGAGACCTCCCA 284
QY 425 GATATGAGCCCACTCTTGTATACGAGCCCACTGAGATATGAGCCCACTC 484
DB 283 GTTCCGAGAGACTCTCTCCAGTTCCGAGAGACTCTCTCCAGTTCCGAGAGACTCTC 224
QY 485 TAGATATGAGCCCACTCTTGTATATGAGCCCACTCTGATATGAGCCCACTC 544
DB 223 CCAAGTCCGAGAGACTCTCAAGTCTGAGAGAGACTCTCCAGTTCTGAGAGAGAC 164
QY 545 CTCTGGATATGAGCCCACTGAGAGAAATGAAGCCCGCTGTGGAGATACAGAGCT 604
DB 163 CTCCAGTTCCGAGAGAACTCTCCAGTTCTGAGAGAGACTCTCCAGTTCCGAGAGAG 104
QY 605 CACCTGTGATCAGAGCCAGGCTC 631

DB 103 CTCCTCCAGTTCCGAGAGACTCTC 77

RESULT 14
US-09-918-995-33536

/ Sequence 33536, Application US/09918995
/ Publication No. US2003007623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES

FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995

PRIOR FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 33536

LENGTH: 432

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

LOCATION: (1) ... (432)

OTHER INFORMATION: n = A,T,C or G

US-09-918-995-33536

Query Match 7.7%; Score 77.2; DB 11; Length 432;

Best Local Similarity 62.4%; Pred.No. 1.2e-14; Matches 121; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 ATGCCATTGATCTGATGACCACTCACTTTAACAACCAAGTATTTGTCGAATTC 60
DB 231 ATGCCATTGATCTGATGACCACTCACTTTAACAACCAAGTATTTGTCGAATTC 290
QY 61 ATTAAAGGAATATTCAGGAGCTCCATATGCTGTGGAGAGCAAGTACTTTTAA 120
DB 291 ATCAAGGAAAGTGAAGCGGAGCGGAGAGTGTGGAGAGCTGTGCTTCAAG 350
QY 121 TTAGCTTCAGAAATGAGATGCCATTTGATTTGCCAGTTAGTGAAGCTGCTCT 180
DB 351 TTGACTTTCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 410
QY 181 GCTGTGCGCAGG 194
DB 411 CAAGCTCCAGAGG 424

RESULT 15
US-10-029-386-11265
/ Sequence 11265, Application US/10029386
/ Publication No. US2003019470A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: ABOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11265
LENGTH: 541
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO Z97205.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
OTHER INFORMATION: SWISSPROT HIT: P22793, EVALUE 1.00e-04

US-10-029-386-11265

Query Match 6.9%; Score 69.2; DB 13; Length 541;
Best Local Similarity 54.3%; Pred. No. 6.1e-12;

Matches 140; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

```
QY 308 CCCACCTGAGATATGAGAGCCCACTCCGATACGAGAGCCCACTGACAGATATG 367
    |||||
Db 14 CTCACCTGAGAGCTCTCTCTCACTGAGAGCTCTCTCTCACTGAGAGCTCTT 73
    |||||
QY 368 GAGCCCAACCGTAGAATAAGAGCCGCTGTGATACAGAGCCTCACTGTGAT 427
    |||||
Db 74 CTCCTGACCTGAGAGGCTCTCTCTCACTGAGAGCTCTCTCTCACTGAGAGCT 133
    |||||
QY 428 ATGAGAGCCCACTCTTGATACGAGAGCCCACTGACAGATATGAGAGCCCACTTAG 487
    |||||
Db 134 CTCTCTCTCACTGAGAGCTCTCTCTCACTGAGAGCTCTCTCTCACTGAG 193
    |||||
QY 488 GATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGATATGAGAGCCCACTC 547
    |||||
Db 194 GCTCCCTCTCTCACTGAGAGCTCTCTCTCACTGAGAGCTCTCTCTCACTG 253
    |||||
QY 548 TCGATATGAGAGCCCACT 565
    |||||
Db 254 CAGGCTCTCTGAGAGCC 271
    |||||
```

Search completed: December 16, 2003, 17:58:45
Job time : 331.43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:21:45 ; Search time 2060.06 Seconds

(without alignments)
11809.735 Million cell updates/sec

Title: US-09-864-291-11

Sequence: 1 atgcctcattgacatgcagac.....atracctcattgcagtag 1001

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_esthum:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_huv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pig:*
27: em_gss_vit:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	699.4	69.9	1079	12	BM564167 AGENCOURT
2	368.2	36.8	663	10	BG699398 602679081
3	362.8	36.2	775	10	BG722815 602695278
4	333.2	33.3	1414	11	AK015863 Mus muscu

Result No.	Score	Query Match	Length	DB ID	Description
5	307.8	30.7	431	9	AM169980
6	271.8	27.2	394	9	AI220013
7	263	26.3	701	10	BG701881
8	252.2	25.2	717	14	BY175414
9	247.8	24.8	379	9	AA812559
10	222.2	22.2	814	13	BU961805
11	202.8	20.3	334	9	AA812560
12	187.2	18.7	332	9	AM104932
13	160.4	16.0	488	28	AQ997933
14	138.8	13.9	426	28	AQ997939
15	129.2	12.9	536	12	BM969697
16	125	12.5	403	9	AA468891
17	118.2	11.8	671	29	AG062842
18	105	10.5	632	28	AZ896232
19	104.2	10.4	444	10	BF000754
20	101	10.1	488	10	BF476944
21	100.6	10.0	568	12	BI390687
22	100.6	10.0	636	12	BM490729
23	98	9.8	499	13	BO356763
24	97.4	9.7	560	13	BU103740
25	97.4	9.7	700	13	BU284977
26	97.4	9.7	723	9	AJ452591
27	97.4	9.7	744	13	BU230211
28	97.4	9.7	932	13	BU107273
29	90.4	9.0	706	13	BU433499
30	89.8	8.9	723	9	AV899050
31	88.8	8.9	755	13	BM089605
32	88.4	8.8	772	9	AJ454950
33	88	8.8	611	9	AL846063
34	87.6	8.8	686	12	BP021598
35	87.2	8.7	594	12	BM190972
36	87.2	8.7	629	10	BR828725
37	87.2	8.7	816	13	BU116890
38	87.2	8.7	931	13	BU914288
39	86.8	8.7	562	12	BM495739
40	86.6	8.7	622	10	BR468886
41	86.6	8.7	899	10	BR470466
42	86.4	8.6	613	9	AL849211
43	86.4	8.6	635	9	AL889954
44	86.4	8.6	647	9	AL849651
45	86.4	8.6	655	9	AL898723

ALIGNMENTS

RESULT 1
LOCUS BM564167
DEFINITION AGENCOURT_6560149 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742129
ACCESSION BM564167
VERSION BM564167.1 GI:18811738
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 1079)
AUTHORS Mamalia; Butcheria; Primates; Catarrhini; Homiidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
plate: LHAM2759 row: m column: 10
High quality sequence stop: 726.

FEATURES

Location/Qualifiers
1. 1079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5742129"
/tissue_type="medulla"
/lab_host="DH10B"
/clone_lib="NIH_MGC_119"
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI;
Site 2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 013. Note:
this is a NIH MGC Library."

BASE COUNT 267 a 297 c 268 g 247 t
ORIGIN

Query Match 69.9%; Score 699.4; DB 12; Length 1079;
Best Local Similarity 96.1%; Pred. No. 6,7e-189;
Matches 749; Conservative 0; Mismatches 26; Indels 4; Gaps 3;

```

OY 1 ATGCCATTGATCTGATGACGAACTCACTGTTGAACAACGATTTTGTGCAAACTTC 60
DB 269 ATGCCATTGATCTGATGACGAACTCACTGTTGAACAACGATTTTGTGCAAACTTC 328
OY 61 ATTAGGGGAACATTCAGGAGCTCCATTTGGTGGTGGGAAGCAAGCTACTTTTAA 120
DB 339 ATTAGGGGAACATTCAGGAGCTCCATTTGGTGGTGGGAAGCAAGCTACTTTTAA 388
OY 121 TTAGCTTCAGAAATGAGATGCCATTTGATTTGCCAGTTGATGATGATGATGATGAT 180
DB 389 TTAGCTTCAGAAATGAGATGCCATTTGATTTGCCAGTTGATGATGATGATGATGATGAT 448
OY 181 GCTGTGGCCGAGGATTTTCACCTTAAATGATGATGATGATGATGATGATGATGATGAT 240
DB 449 GCTGTGGCCGAGGATTTTCACCTTAAATGATGATGATGATGATGATGATGATGATGAT 508
OY 241 TATGTAATTAATCTGGGGAAGGGAATATGTCATCCACAGATGCTTTGATTAATGTC 300
DB 509 TATGTAATTAATCTGGGGAAGGGAATATGTCATCCACAGATGCTTTGATTAATGTC 568
OY 301 TATGAGGCCCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 360
DB 569 TATGAGGCCCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 628
OY 361 GGATATGAGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 420
DB 629 GGATATGAGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 688
OY 421 GTGCAATATGAGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 479
DB 689 GTGCAATATGAGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 748
OY 480 ACCTCTAGGAATATGAGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 539
DB 749 ACCTCTAGGAATATGAGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 808
OY 540 CCCACCTCTCGGATATGAGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 599
DB 809 CCCACCTCTCGGATATGAGCCCACTGCAAGATATGAGCCCACTCCCGATACGAGCCCACTGCA 868
OY 600 AGCCATCACTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 657
DB 869 AGCCATCACTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 928
OY 658 GAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 717
DB 929 GAAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 988

```

OY 718 ATGTAA-CTTGAAGACTCACAGAAAGGATGATGATGATGATGATGATGATGATGATGAT 775
DB 989 ATGTAACTTGAAGACTCACAGAAAGGATGATGATGATGATGATGATGATGATGATGAT 1047

RESULT 2
BG699398
LOCUS
DEFINITION
602679081P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5',
mRNA sequence.
ACCESSION
BG699398
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 663)
NIH-MGC <http://mgs.nci.nih.gov/>.
NATIONAL INSTITUTES OF HEALTH, MAMMALIAN GENE COLLECTION (MGC)
JOURNAL
COMMENT
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
plate: LLM10703 row: a column: 21
High quality sequence stop: 660.

FEATURES
source
Location/Qualifiers
1. 663
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4811804"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: Bluescript (modified
pBluescript KS+); Site: 1: BamHI; Site 2: SalI-XhoI (gagag
); Oligo-dT primed using primer 5'-TTTTTATTTTATTTTAA-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 168 a 158 c 169 g 164 t
ORIGIN

Query Match 36.8%; Score 368.2; DB 10; Length 663;
Best Local Similarity 97.1%; Pred. No. 4e-94;
Matches 395; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

```

OY 1 ATGCCATTGATCTGATGACGAACTCACTGTTGAACAACGATTTTGTGCAAACTTC 60
DB 268 ATGCCATTGATCTGATGACGAACTCACTGTTGAACAACGATTTTGTGCAAACTTC 317
OY 61 ATTAGGGGAACATTCAGGAGCTCCATATGATGATGATGATGATGATGATGATGATGAT 120
DB 318 ATTAGGGGAACATTCAGGAGCTCCATATGATGATGATGATGATGATGATGATGATGAT 377
OY 121 TTAGCTTCAGAAATGAGATGCCATTTGATTTGCCAGTTGATGATGATGATGATGATGAT 180
DB 378 TTAGCTTCAGAAATGAGATGCCATTTGATTTGCCAGTTGATGATGATGATGATGATGAT 437
OY 181 GCTGTGGCCGAGGATTTTCACCTTAAATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 438 GCTGTGGCCGAGGATTTTCACCTTAAATGATGATGATGATGATGATGATGATGATGATGAT 497

```

QY	241	TATGAAATTA	CTGGGGGA	GGGAAATAT	TGCACTC	CAAGAG	CCCTGT	GTGCAATAT	ATGTC	300
DB	498	TATGAAATTA <th>CTGGGGGA</th> <th>GGGAAATAT</th> <th>TGCACTC</th> <th>CAAGAG</th> <th>CCCTGT</th> <th>GTGCAATAT</th> <th>ATGTC</th> <td>557</td>	CTGGGGGA	GGGAAATAT	TGCACTC	CAAGAG	CCCTGT	GTGCAATAT	ATGTC	557
QY	301	TATGGGGCCCA <th>CTCGAGAT</th> <th>TATGAGAC</th> <th>CCCACTC</th> <th>CCGATAC</th> <th>CGAGCCCA</th> <th>CTGCA</th> <th>360</th>	CTCGAGAT	TATGAGAC	CCCACTC	CCGATAC	CGAGCCCA	CTGCA	360	
DB	558	TATGAGCCCA <th>CTCGAGAT</th> <th>TATGAGAC</th> <th>CCCACTC</th> <th>CCGATAC</th> <th>CGAGCCCA</th> <th>CTGCA</th> <th>615</th>	CTCGAGAT	TATGAGAC	CCCACTC	CCGATAC	CGAGCCCA	CTGCA	615	
QY	361	GGATTTGAGC <th>CCCACTCG</th> <th>TAGAAATGA</th> <th>AGCCCG</th> <th>CCTGTG</th> <th>GAGTA</th> <th>407</th>	CCCACTCG	TAGAAATGA	AGCCCG	CCTGTG	GAGTA	407		
DB	616	GGATTTGAGC <th>CCCACTCG</th> <th>TAGAAATGA</th> <th>AGCCCG</th> <th>CCTGTG</th> <th>GAGTA</th> <th>662</th>	CCCACTCG	TAGAAATGA	AGCCCG	CCTGTG	GAGTA	662		
RESULT 3										
LOCUS	Bg722815									
DEFINITION	6026952781 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5',									
ACCESSION	Bg722815									
VERSION	Bg722815.1									
KEYWORDS	EST.									
SOURCE	Homo sapiens (human)									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;									
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.									
TITLE	1 (bases 1 to 775)									
JOURNAL	NIH-MGC http://mgs.ncl.nih.gov/ .									
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished									
	Contact: Robert Strausberg, Ph.D.									
	Email: cgapbs-remail.nih.gov									
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.									
	cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki									
	Toshiyuki and Piero Carninci (RIKEN)									
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)									
	DNA Sequencing by: Incyte Genomics, Inc.									
	Clone distribution: MGC clone distribution information can be									
	found through the I.M.A.G.E. Consortium/LNLN at:									
	http://image.lnl.gov									
	Plate: LLM10743 row: 1 column: 20									
	High quality sequence stop: 651.									
FEATURES										
source										
	1..775									
	/organism="Homo sapiens"									
	/mol_type="mRNA"									
	/db_xref="taxon:9606"									
	/clone="IMAGE:4827427"									
	/lab_host="DH10B"									
	/clone_11b="NIH MGC 97"									
	/note="Ribbon: testis; Vector: BluescriptR (modified									
	pBluescript KS+), Site_1: BamHI, Site_2: SalI-XhoI (gcgcgag									
) ; Oligo-dT primed using primer 5'-ttttttttttttttt-3',									
	size-selected for average insert size 2.2 kb and									
	normalized to ROT 5. This is a primary library enriched									
	for full-length clones and constructed using the									
	Cap-trapper method (Carninci, in preparation). Library									
	constructed by M. Brownstein (NIH/NHGRI, National									
	Institutes of Health). Note: this is a NIH_MGC Library."									
BASE COUNT	191 a 212 c 193 g 179 t									
ORIGIN										
Query Match	36.2%	Score 362.8;								

QY	12	TTATCTCTTCAGAAATGAGATGCGATTTGAATTTGGCCAGTTGATGGAAGAGCTCT	180
Db	396	TTATCTCTTCAGAAATGAGATGCGATTTGAATTTGGCCAGTTGATGGAAGAGCTCT	455
QY	181	GCTGTTGCCAGAGATTTCCACTTAGAACCTTAA--TGACTGTTGAGCTCTTAGGGAAT	239
Db	456	GCTGTTGCCAGAGATTTCCACTTAGAACCTTAA--TGACTGTTGAGCTCTTAGGGAAT	515
QY	240	TTATGTAATTACTGAGGAGAGGAATATATGCACTCCACAGATGCTTGTCACTTATGT	299
Db	516	TTATGTAATTACTGAGGAGAGGAATATATGCACTCCACAGATGCTTGTCACTTATGT	575
QY	300	CTATGAGGAGCCCACTGTCAGAG--ATATGAGAGCCCACTCCCGATATAGAGCCCACTG	358
Db	576	CTATGAGAGCCCACTGTCAGAGCATATGAGAGCCCACTCCCGATATAGAGAG--CCAACTG	634
QY	359	CAGGATATGAGAGCCCACTGTCAGAGATATAGAGCCCGCTGTCAGATATAGAGCCCTC-A	417
Db	635	CAGGATATGAGAGCCCACTGTCAGAGATATAGAGAGCCCGCTGTCAGATATAGAGCCCTCA	694
QY	418	CTGTGTGGA-TATGAGAGCCCACTCTTGG	446
Db	695	CTGTGTGGA-TATGAGAGCCCACTCTTGG	724
RESULT 4			
LOCUS	AK015863	1414 bp	mRNA
DEFINITION	AK015863	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930521123 product:hypothetical protein, full insert sequence.	
ACCESSION	AK015863	GI:12854363	
KEYWORDS	HTC; CAP trapper.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.		
REFERENCE			
AUTHORS	1 Carninci, P. and Hayashizaki, Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)		
MEDLINE	99279253		
PUBMED	10349636		
REFERENCE	2		
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes		
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)		
MEDLINE	20499374		
PUBMED	11042159		
REFERENCE	3		
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, D., Nishi, K., Kitanai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishize, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohtsuka, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.		
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer		
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)		
MEDLINE	20530913		
PUBMED	11076861		
REFERENCE	4		
AUTHORS	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Komno, H., Aichi, J., Fukuda, S., Aikawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadote, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Giesi, C., King, B., Kochiwa, H.,		

ACCESSION	AM169980
VERSION	AM169980.1
KEYWORDS	GI:6401505
SOURCE	EST.
ORGANISM	Homo sapiens (human)
	Homosapiens
	Mammalia; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteleostomi; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	1 (bases 1 to 431) NCI-CCAF http://www.ncbi.nlm.nih.gov/ncicgap .
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgsap@remail.nih.gov This clone is available royalty-free through LNL ; contact the IMAG Consortium (infoimage.jnl.gov) for further information. Seq primer: -40UP from Glibco High quality sequence stop: 426.
FEATURES	location/Qualifiers
source	1..431
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="IMAGE:2659201"
	/lab_host="DH10B"
	/clone_lib="Soares_NFL_T_GBC_S1"
	/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalised libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI CGAP GCBI) were mixed, and 88 circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The divergent PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.B. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT	114 a 107 c 82 g 127 t 1 others
ORIGIN	
Query Match	30.7%; Score 307.8; DB 9; Length 431;
Best Local Similarity	99.0%; Pred.No.6.1e-77;
Matches 309; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
Dn	690 TTCTCAGGTCCATTCTTAACCTTTCAAGATGTAACCTTGAAAGCTCAACAAGAAG 749
	431 TTCTCAGGTCCATTCTTAACCTTTCAAGATGTAACCTTGAAAGCTCAACAAGAAG 372
Dy	750 GTACCCTAATAATGAAGTCAGGATTAAGAGAGAGCATCAGCTTAGAGTCATTGATTAATC 809
	371 GTACCCTAATAATGAAGTCAGGATTAAGAGAGAGCATCAGCTTAGAGTCATTGATTAATC 312
Dz	810 TGCAATGGAAAATTGAGAAACCAATGCTGCCAGTTCTCAAGAACCAACTTTCTTAAG 869
	311 TGCAATGGAAAATTGAGAAACCAATGCTGCCAGTTCTCAAGAACCAACTTTCTTAAG 252
Dc	870 GAAGTCAGTACGTGGCAAACAGTGAATGAGAGAGAGAGAGAGACTCAACCACTAGAG 929
	251 GAAGTCAGTACGTGGCAAACAGTGAATGAGAGAGAGAGAGAGACTCAACCACTAGAG 192
Dd	930 CAGGATTAAGTTTCCCTTTGTTCACTTTTCAGTGTCTGCTGGAAATGATGATTAACCTC 989
	191 CAGGATTAAGTTTCCCTTTGTTCACTTTTCAGTGTCTGCTGGAAATGATGATTAACCTC 132
Dg	990 ATTGTCAGGTAG 1001
	131 ATTGTCAGGTAG 120
RESULT 6	
AT1220013/c	A1220013
LOCUS	394 bp mRNA linear EST 30-NOV-1996

DEFINITION	cg78ell.xl Soares NPL T GBC S1 Homo sapiens cDNA clone					
ACCESSION	IMAGE:1841324 3', mRNA sequence.					
VERSION	A1220013					
KEYWORDS	A1220013.1	GI:3802216				
SOURCE	BST.					
ORGANISM	Homo sapiens (human)					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Butioria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1 (bases 1 to 394)					
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.					
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),					
JOURNAL	Tumor Gene Index					
COMMENT	Unpublished					
	Contact: Robert Strausberg, Ph.D.					
	Email: cgaps-remail.nih.gov					
	This clone is available royalty-free through LML ; contact the					
	IMRG Consortium (info@image.llnl.gov) for further information.					
	Insert Length: 1264 Std Error: 0.00					
	Seq primer: -40UP from Gdbco					
	High quality sequence stop: 368.					
FEATURES	Location/Qualifiers					
source	1..394					
	/organism="Homo sapiens"					
	/mol_type="mRNA"					
	/db_xref="taxon:9606"					
	/clone="IMAGE:1841324"					
	/lab_host="DH10B"					
	/clone_id="Soares NPL T GBC S1"					
	/notes="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with					
	a modified polylinker; Site 1: Not I; Site 2: Eco RI;					
	Equal amounts of plasmid DNA from three normalized					
	libraries (fetal lung NBH149w, testis NT, and B-cell					
	NCI CGAP GCBI) were mixed, and as circles were made in					
	vitro. Following HAP purification, this DNA was used as					
	tracer in a subtractive hybridization reaction. The driver					
	was PCR-amplified cDNAs from pools of 5,000 clones made					
	from the same 3 libraries. The pools consisted of					
	I.M.A.G.E. clones 297480-302087, 682632-687239,					
	726408-728711, and 729096-731399. Subtraction by Bento					
	Soares and M. Fatima Bonaldo. "					
BASE COUNT	101 a 103 c 73 g 117 t					
ORIGIN						
Query Match	27.2%; Score 271.8; DB 9; Length 394;					
Best Local Similarity	99.3%; Pred. No. 1.2e-66;					
Matches 273; Conservative % 0; Mismatches 2; Indels 0; Gaps 0;						
CY	727 TTGAAGACTCACCAAGAAAGGTACCCTTAATTGAAGTCAGATTAAGAAGACGACT	786				
DB	394 TTGAAGACTCACCAAGAAAGGTACCCTTAATTGAAGTCAGATTAAGAAGACGACT	335				
CY	787 CAGCTTAGAGCATTTAATGATTCGATTTGGAAAATTAGAAACGATGCTCCAGTG	846				
DB	334 CAGCTTAGAGCATTTAATGATTCGATTTGGAAAATTAGAAACGATGCTCCAGTG	275				
CY	847 TCTCAAGACGACCTTTCTTAAGGAAGTCAGTACGTGGCAACAGTATAGAGAAAG	906				
DB	274 TCTCAAGACGACCTTTCTTAAGGAAGTCAGTACGTGGCAACAGTATAGAGAAAG	215				
CY	907 AGAGAGACTAACCACTAGACGAGGATTAAGTTCCCTTTGCAGCTTTTCAGTGC	966				
DB	214 AGAGAGACTAACCACTAGACGAGGATTAAGTTCCCTTTGCAGCTTTTCAGTGC	155				
CY	967 TGCTGAATGATGATTAATCCTCATTTGCAAGTGA 1001					
DB	154 TGCTGAATGATGATTAATCCTCATTTGCAAGTGA 120					
RESULT 7						
LOCUS	BC701881	701 bp	mRNA	linear	EST 07-MAY-2001	
DEFINITION	60268323F7L NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4815849 5' ,					

mRNA sequence.
 BG701881 GI:13972665
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Bacteria; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NIH-MGC http://mgs.nci.nih.gov/
 1 (bases 1 to 701)
 REFERENCE
 AUTHOR S
 TITLE
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.llnl.gov
 Plate: L1AM10713 row: j column: 10
 High quality sequence stop: 680.
 Location/Qualifiers
 1..701
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4815848"
 /tissue_type="hippocampus"
 /lab_host="DH10B"
 /clone_id="NIH_MGC_95"
 /note="Organ: brain; Vector: pBluescriptpr (modified
 pBluescript KS+); Site 1: BamHI, Site 2: SalI-XhoI (gcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
 size selected for average insert size 2.5 kb and
 normalized to 10⁷ c. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC library."
 BASE COUNT
 176 a 164 c 189 g 172 t
 ORIGIN
 Query Match 26.3%; Score 263; DB 10; Length 701;
 Best Local Similarity 89.4%; Pred. No. 5.9e-64;
 Matches 387; Conservative 0; Mismatches 30; Indels 16; Gaps 9;
 QY 1 ATGCCATTGATCTGATGACGAACCTC--ACTGTTGAACACCACTATTGCT--GCAA 55
 DB 261 ATGCCATTGATCTGATGACGAACCTC--ACTGTTGAACACCACTATTGCTGCGGAA 320
 QY 56 ACTTCATTAAAGGAACCTATTCAGGACGCTCCATATGCTGCTGGGAGGAGCAAGTACT 115
 DB 321 CTTCGATTAAAGGAACCTATTCAGGACGCTCCATATGCTGCTGGGAGGAGCAAGTACT 380
 QY 116 TT---AAATTAGTCTGAGAAATGAGATGTCATTGAATTGCGCAGTGTATGCTGAA 171
 DB 381 TTATTAATGTAAGTCTTCAGAAATGAGATGTCATTGAATTGCGCAGTGTATGCTGAA 440
 QY 172 GCTGCTCTGCTGTTCCCGAGAGATTTCCACTTAGAACCTTAA-ATGACTGTTAGCTC 230
 DB 441 GCGGCTCTGCTGCTGCGCCGAGAGATTTCCACTTAGAACCTTAAATGACTGTTAGCTC 500
 QY 231 TATGGGAATTATGTAATTAATCTGCGGAGGAGAAATATGCA--CTCCACAGATGCTTGT 288
 DB 501 TATGGGCAATTATGTAATTAATCTGCGGAGGAGAAATATGCAATGTCAGTCTTGT 560
 QY 289 TGAAGTATGCTATGAGGAGGAGCCCACTGACAGATATGAGGAGCCCACTCCGATACGGA 348
 DB 561 ACGATTATGCTATGAGGAGGAGCCCACTGACAGATATGAGGAGCCCACTCCGATACGGA 618

QY 349 GCCCACTCTGACGATATGAGGCCCAACCCGTAGAAATGAGGCCGCTGTGGATAC 408
 DB 619 GCCCACTCTGACGATATGAGGCCCAACCCGTAGAAATGAGGCCGCTGTGGATAC 676
 QY 409 AGAGCTCTACCTG 421
 DB 677 AGAGCTCTACCTG 689
 RESULT 8
 BY715414
 LOCUS
 DEFINITION
 BY715414 RIKEN full-length enriched, adult male testis Mus musculus
 cDNA clone 493052123 5', mRNA sequence.
 BY715414
 BY715414.1 GI:27128531
 EST.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus (house mouse)
 Mus musculus
 Bacteria; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 717)
 REFERENCE
 AUTHORS
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,
 Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batilov, S.,
 Beisel, K. W., Blake, J. A., Brad, D., Brusic, V., Chochia, C., Corbani,
 L. B., Cousins, S., Dalla, B., Dragani, T. A., Fletcher, C. F., Forrest,
 A., Fraser, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, J. J.,
 Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kozlowski, R. M.,
 King, B. L., Konegaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,
 P. A., Meglath, D. R., Maltse, L., Marchionni, L., McKenzie, L., Miki,
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Petrea, G.,
 Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D.,
 Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,
 B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sample, C. A., Setou,
 M., Shinada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,
 R. D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y.,
 Watanabe, Y., Wells, C., Wilming, L. G., Wysshaw-Boris, A., Yangisawa,
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
 Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
 K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
 R. S., Rogers, J., Birney, B. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 TITL
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda,
 S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,
 Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno,
 H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K.,
 Nunazaki, R., Ohno, M., Osato, N., Saito, R., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tegan, M., Takeda, Y.,
 Waki, K., Watanabe, A., Watanabe, M. and Hayashizaki, Y. Direct
 Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multipillar sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedias: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
 Location/Qualifiers

FEATURES

SOURCE

1..717
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="4930521123"
 /sex="male"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="SOLR"
 /clone_lib="RIKEN full-length enriched, adult male testis"
 /note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'-GAGGAGAGAGATCCAGACGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence [5'-GAGGAGAGAGCGGCGGCAATTAATCTCGAGTTAATTAATCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites."

BASE COUNT

161 a 193 c 198 g 165 t

ORIGIN

Query Match 25.2%; Score 252.2; DB 14; Length 717;
 Best Local Similarity 74.5%; Pred. No. 7.5e-61;
 Matches 333; Conservative 0; Mismatches 108; Indels 6; Gaps 1;

1 ATGCCATTGATCGAGACGACCTCTGTTGAACACGAGTATTCGCAAACTTC 60
 277 ATGCCATTGATCGAGACGACCTCTGTTGAACACGAGTATTCGCAAACTTC 336
 61 ATTAAGGAACTATTCAGGAGCTCCATATGATGCTGGAGGACAAAGTACTTTTAA 120
 337 ATTAAGGAACTATTCAGGAGCTCCATATGATGCTGGAGGACAAAGTACTTTTAA 396
 121 TTAGCTTCAGAAATGAGATGCCATTGAATTTGCCAGTATGATGTAAGCTGCTCT 180
 397 ATAGCTTCCTCCGAAAGAGAGTGTATGACTTTGGCCAGTATGAGCCAAAGCTGCTCT 456
 181 GCTGTGGCCGAGATTTCCACTTAAGACCTTAATGACGTGTCAGCTCATAGGAATT 240
 457 GCGGCTGCCAAGAGATTCCTACGAGTTGCAAGCTTCTGATAGGCGCTCTGAGAAAT 516
 241 TATGTAATTAATTCGAGGAGAAATATGCACTCCACAGATGCTGTTCTAGTTATGTC 300
 517 TATGTCATCACTGGGAGACGAACTATATGCCCAAG-----GCATACCAAGTTGCC 570
 301 TATGGAGCCCACTGACAGATATGAGACCCCACTCCCGAGTACGAGGCCCACTTCA 360
 571 TATGAGCCCACTGACAGATATGAGACCCCACTCCCGAGTACGAGGCCCACTTCA 630
 361 GATATGAGCCCACTGAGAAATGAGGCCCGCTGTGGGATACAGAGCTCACT 420

Db 631 GGGTATGAGACCCCACTGAGGAGTACGAGCCCGCGTGGTATATTAACCACTT 690
 Oy 421 GTGCAATGAGACCCCACTTGTGA 447
 Db 691 CCCGATACGATGCTGCTCTGGA 717

RESULT 9
 AA812559/c 379 bp mRNA linear EST 31-DEC-1998
 LOCUS AA812559.1
 DEFINITION a31c01.81 Soares_testis_NHT Homo sapiens cDNA clone 1391904 3', mRNA sequence.
 ACCESSION AA812559
 VERSION AA812559.1 GI:2882623
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 379)
 NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 COMMENT Email: CGAP@remail.nih.gov
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: www.bio.lnl.gov/bbrp/image/image.html
 Insert Length: 1331 Std Error: 0.00
 Seq primer: -40ml3 fwd. RT from Amerham
 High quality sequence stop: 311.
 Location/Qualifiers

FEATURES

SOURCE

1..379
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="1391904"
 /sex="male"
 /lab_host="DH10B"
 /clone_lib="Soares_testis_NHT"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5'-TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cos5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

98 a 97 c 68 g 116 t

ORIGIN

Query Match 24.8%; Score 247.8; DB 9; Length 379;
 Best Local Similarity 99.2%; Pred. No. 8.8e-60;
 Matches 249; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

751 TACCTAAATGAGTGCAGATAGAGAGACGACTGAGTTAGATGATTAATCT 810
 Db 379 TACCTAAATGAGTGCAGATAGAGAGACGACTGAGTTAGATGATTAATCT 320
 Oy 811 GCATTGGAATAATAGAAACCGAGTCCCATGTTCTCAAGACAGCCCTTTTAAG 870
 Db 319 GCATTGGAATAATAGAAACCGAGTCCCATGTTCTCAAGACAGCCCTTTTAAG 260
 Oy 871 AAGTCATGCTGGGCAACAGTATGAGAGAGAGAGAGACTCAACCACTAGAGC 930

Db	259	AAGTCACGACTGGGCACAATGATGAGAAGAGAGAACTCAACCAACTAAGC	200
OY	931	AGGATAAGTTCCCTTGTTCAGCTTTCACTGTCTCTCGAATGATATTACTTCA	990
Db	199	AGGATAAGTTCCCTTGTTCAGCTTTCACTGTCTCTCGAATGATATTACTTCA	140
OY	991	TTCACAGTAG 1001	
Db	139	TTCACAGTAG 129	
RESULT 10			
BU961805		814 bp mRNA linear EST 21-OCT-2002	
LOCUS			
DEFINITION	AGNCOURT 10617160 NIH MGC 169 Mus musculus cDNA clone		
FEATURES	IMAGE:674Z447 5', mRNA sequence.		
ACCESSION	BU961805		
VERSION	BU961805.1 GI:24191377		
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 814)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished		
COMMENT	Contact: Robert Strauberg, Ph.D.		

Tissue Procurement: Dr. Jonathan Kuo, NIMH
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLCM3080 row: 1 column: 06
High quality sequence stop: 504.
Location/Qualifiers
1. .814

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:574247"
/lab_host="DH10B (T1-phage-resistant)"
/clone_1ib="NIH_MGC_169"
/notes="Organ: Testicles; Vector: pDNR-LIB; Site_1: SfiI
(ggcattatggcc); Site_2: SfiI (ggcggccctggcc); cDNA made
by oligo-dt priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAGCCAGAGTGGCCATTACGCCGGG-3' and
5'-ATTCTAGAGCGCGAGCGGCCGACATG-dt(30)NN-3'. P111-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in laboratory of M.
Brownstein (NIH). Note: this is a NIH_MGC Library."

```

Query Match	22.2%	Score 222.2	DB 13	Length 814
Best Local Similarity	67.2%	Pred. No. 3.2e-52		
Matches 314	Conservative 0	Mismatches 153	Indels 0	Gaps 0
QY	325	GGAGCCCACTCCCGATACGAGGCCCACTGCAGATATGAGCCCAACCGTAGGA	384	
Db	3	GGGTCCCTCCTCTCATATGTAGTACCCCAATGGGCTATGAGAGTTCACCTCCGGA	62	
QY	385	AATGAAGGCCCGCGCTGTGGGATACAGACCTGACCTGTGGAGATATGAGGCCCACTCTT	444	
Db	63	TATGGACCTTCACCTGTGAGATAGGATATCCCACTCTTGATATGAGGCCCACTATG	122	
QY	445	GGATAGCGAGCCCACTGCAGATATGAGGCCCACTTAGATATGAGGCCCACT	504	

Db	123	GAGTATGAGGCCAACTCCCAAGTACGGAACTACCTTGGAGTCCGGTCTCCACT	182
Qy	505	CTTGGATTGGAACCCCACTCTCCGATATGAGGCCCACTCTCAGGATATGAGGCCCA	564
Db	183	CCAGATTATGAGCCCCGCTTATGGGGTATGAACTCTCTCTTCAGGGCGTTAGTCCATA	242
Qy	565	CCTGCAGGAATGAAAGGCCCGCTTCGCGGAATACAGAGCTCACTGCTGATCAGAGCC	624
Db	243	CCTCTCGATCAAGGGCAAGTCTGTAGCAACAGAACTCACTGCTGGGTCTGAAGCA	302
Qy	625	AGGCTCAGGAATCTACAGAGAGCCAGAGCTCTGAAAAAGAGCTTCTTCCCTGACC	684
Db	303	GGCATCCCAATCTGTGGCAGATCCAAATCTGAAATTCAGGAGATCTTTTCCCTAAC	362
Qy	685	TCTCTTCTCAGGTCCATCTTAACTTCTTAAGATGAAACCTTGAAGCTCACCAAGCA	744
Db	363	TGTGTCCTCAAGTCCACTCTCCCGCTCTAAGATGAAACCTTGAAGTTTCACCAAGCA	422
Qy	745	AAAGAGTACCTTAATTTGAAGTCAGAGTAAAGAGACGACTCAGCT	791
Db	423	AAACGACCTTAATGTTGAAGTCAGATTAACCGGAACCTTAGGT	469

RESULT 11
AA812560/c 334 bp mRNA linear EST 31-DEC-1998
LOCUS a11c02.s1 Soares_testis_NHT Homo sapiens cDNA clone 1391906 3',
DEFINITION mRNA sequence.
ACCESSION AA812560
VERSION AA812560
KEYWORDS GI:2882624
SOURCE EST.
ORGANISM Homo sapiens (human)
Homologs Homo sapiens
Eukaryotes Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 334)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonald
, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMDL at:
www-bio.lnlnl.gov/bbrp/image/image.html
Insert length: 1310 Std Error: 0.00
Seq primer: -40m13 fwd. Rr from Amerham
High quality sequence stop: 189.
Location/Qualifiers
1. .334

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="1391906"
/sex="male"
/lab_host="DH10B"
/clone_1lb="Soares_testis_MHT"
/notes="Vector: pTVT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clonech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGCCGCCCATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTVT3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

```

	Query Match	Best local	Similarity	Score	202.8	DB	9	Length	334
	Matches	204	Conservative	0	Mismatches	2	Indels	0	Gaps
QY	796	GTCAATTGATGATCGATTTGAAAATTAAGAAACCGATGCTCCAGTGTCTCAAGA	855						
Db	334	GTCAATTGATGATCGATTTGAAAATTAAGAAACCGATGCTCCAGTGTCTCAAGA	275						
QY	856	CGACTTCTTAAAGAAAGTCAGTACGTGGGCAACAGCATAGAGAAAGAGAGAGAC	915						
Db	274	CAGCTTCTTAAAGAAAGTCAGTACGTGGGCAACAGCATAGAGAAAGAGAGAGAC	215						
QY	916	TCAACCACTAGACAGAGGATTAAGTTCCCTGTTCAGCTTTCAAGTCTCGAGAT	975						
Db	214	TCAACCACTAGACAGAGGATTAAGTTCCCTGTTCAGCTTTCAAGTCTCGAGAT	155						
QY	976	GTGATGATTAAGTCAATTTGCAAGGAG	1001						
Db	154	GTGATGATTAAGTCAATTTGCAAGGAG	129						

RESULT	12
AM104932/c	
LOCUS	AM104932 322 bp mRNA linear EST 20-OCT-1999
DEFINITION	xcd73ad2.x1 Soares_NFL_T_Gbc_S1 Homo sapiens cDNA clone IMAGE:2603210 3', mRNA sequence.
ACCESSION	AM104932
VERSION	AM104932.1 GI:6075667
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE
1 (bases 1 to 322)
AUTHORS
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-r@mail.nih.gov
This clone is available royalty-free through LNL, contact the
IMAG Consortium (info@image.lnl.gov) for further information.
Seq primer: -40UP from glibco
High quality sequence stop: 202.

FEATURES	Location/Qualifiers
source	1. .322

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2603210"
/lab_host="DH10B"
/clone_lib="Soares_NFL_T_GBC_s1"
/notes="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NDH1.9W, testis NHT, and B-cell
NCI CGAP GCB1) were mixed, and ss circles were made
in vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The diverge
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687739,
726408-728711, and 729096-731391. Subtraction by Bento
Soares and M. Fatima Bernaldo."

```

Query Match	18.7%	Score 187.2	DB 9	Length 322
Best Local Similarity	98.4%	Pred. NO. 1.8e-42		
Matches 189	0	Mismatches 3	Indels 0	Gaps 0
Conservative				
810	TGCATTGTGAAATTAAGAAACCGAGATCTCCATGTTCTCAAGACGACCTTTCTTTAAG	869		

Db 322 TGCATTGTGAAAATTAGGAAACGATGCTCCCATGTTCTCAAGSACAGCTTTTAAAG 263
 QY 870 GAAGTCAGTACGTGGGCAACGTGATGAGAGGAAGAGAGAGACTCAACCACTAGAG 929
 Db 262 GAAGTCAGTACGTGGGCAACGTGATGAGAGGAAGAGAGAGACTCAACCACTAGAG 203
 QY 930 CAGGGATTAAGTTCCTTGTTCAGCTTTCACTGTCTGCTGGAATGATGATTAACCTC 989
 Db 202 CAGGGATTAAGTTCCTTGTTCAGCTTTCACTGTCTGCTGGAATGATGATTAACCTC 143
 QY 990 ATTGTCAGTAG 1001
 Db 142 ATTGTCAGTAG 131

RESULT 13	AO997933/c	488 bp	DNA	linear	GSS 24-FEB-2000
LOCUS	AO997933				
DEFINITION	RPCT-23-38420.TJ RPCT-23 Mus musculus genomic clone RPCT-23-38420				
ACCESSION	AO997933				
VERSION	AO997933.1				
KEYWORDS	GSS.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (Pages 1 to 488)	Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Aktinret, B., Levine, M., Mogam, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.	Mouse BAC End Sequences from Library RPCI-23	Unpublished	
Other GSSs:			RPCI-23-384A20.TV	

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (http://www.tigr.org/tcdb/Bac_ends/mouse/bac_end_intro.html)
Plate: 384 row: A column: 20
Seq primer: SP6
Class: BAC ends.

```

FEATURES
source
location/Qualifiers
1..488
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-384A20"
/sex="Female"
/180_host="DH10B"
/clone_1b="RPCI-23"
/notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
EcORI; Site 2: BcORI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcORI and BcORI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT
102 a 128 c 157 g 100 t 1 others
ORIGIN
Query Match 16.0%; Score 160.4; DB 28; Length 488;
Best Local Similarity 67.3%; Pred. No. 1,2e-34;

```

Matches 243; Conservative 0; Mismatches 112; Indels 6; Gaps 1;

```

QY 269 TGAATTAATGTCATGAGGGGCCCCCACTGAGAGATATAGAGCCCACTCCCGAGATACGA 348
DB 457 TCACCGGAGGATACGAGATCCCATCTGAGAGGTATAGAGCCCACTCCCGAGATACGA 398
QY 349 GCGCCACCTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGAGATAC 408
DB 397 GCGCCACCTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGAGATAC 338
QY 409 AGAGCTCAGCTGAGATATGAGAGCCCACTGAGAGATATAGAGCCCACTCCCGAGATAC 468
DB 337 GAGAG-----CTGATGATATGAGCTCCCTCCCTCTATATGATAGCTACCCCAATGAGC 284
QY 469 TATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATAGAGCCCACTCCCGAGATAC 528
DB 283 TATGAGATTCAGCTCTGATATGAGAGCCCACTGAGATATAGAGCCCACTCCCGAGATAC 224
QY 529 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATAGAGCCCACTCCCGAGATAC 588
DB 223 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATAGAGCCCACTCCCGAGATAC 164
QY 589 GCGGATATGAGAGCTGAGATATGAGAGCCCACTGAGATATAGAGCCCACTCCCGAGATAC 648
DB 163 ATGGGATCCGATTCCCACTCCAGATATGAGAGCCCACTGAGATATAGAGCCCACTCCCGAGATAC 104
QY 649 C 649
DB 103 C 103

```

RESULT 14

A0997939/c

LOCUS A0997939 426 bp DNA linear GSS 24-FEB-2000
 DEFINITION RPI-23-384A21.TV RPI-23 Mus musculus genomic clone RPI-23-384A21
 ACCESSION A0997939
 VERSION A0997939.1 GI:7072959

KEYWORDS

GSS.
 Mus musculus (house mouse)

SOURCE

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Zhao, S., Niernan, W., Feldblum, T., Malek, J., Shatsman, S., Akimret,
 B., Levine, M., McGann, S., Tegay, G., Geer, K., Krol, W., de Jong, P.
 and Frazer, C.M.
 Mouse BAC End Sequences from Library RPI-23

TITLE

JOURNAL Unpublished
 COMMENT Other GSS: RPI-23-384A21.TV

Contact: Shaying Zhao
 Department of Bukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

FEATURES

Source Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.bufileo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)
 or from Resea.ch Genetics (info@res.ch). BAC end page:
 http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
 Plate: 384 Row: A Column: 21
 Seq primer: SP6
 Class: BAC ends.
 Location/Qualifiers

1. 426
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPI-23-384A21"

/sex="Female"
 /lab_host="DH10B"
 /clone_lib="RPI-23"
 /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI methylase. Site
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."
 ORIGIN 88 a 107 c 134 g 97 t

Query Match 13.9%; Score 138.8; DB 28; Length 426;
 Best Local Similarity 66.9%; Pred. No. 1.7e-28;
 Matches 230; Conservative 0; Mismatches 107; Indels 7; Gaps 2;

```

QY 307 GCGCCACCTGAGATATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 366
DB 425 GTCCCATCTGAGAGGTATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 366
QY 367 GAGCCCACTGAGAGGTATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 426
DB 365 GTAGCCCATCTGAGAGGTATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 312
QY 427 TATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 486
DB 311 TATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 252
QY 487 GATATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 546
DB 251 GATATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 192
QY 547 CTCGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 605
DB 191 ATGAGATGAGAGCACTGAGATATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 132
QY 606 ACTGCTGATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 649
DB 131 ACTGCTGATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGAGATATAGAGCCCACTCCCGAGATAT 88

```

RESULT 15

BM969697/c

LOCUS BM969697 596 bp mRNA linear EST 20-FEB-2003
 DEFINITION UT-CF-DUI-abc-a-10-0-UT.81 UT-CF-DUI Homo sapiens cDNA clone
 ACCESSION BM969697
 VERSION BM969697.1 GI:19587284

KEYWORDS

EST.
 Homo sapiens (human)

SOURCE

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)

TITLE

JOURNAL MEDLINE
 PUBMED

COMMENT

Contact: McCray, FB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research

Genetics (www.reagen.com) or from Open Biosystems
(www.openbiosystems.com).

The following repetitive elements were found in this cDNA

Sequence: 1-29, >AT-rich#low_complexity (matched complement) 37-73,
>AT-rich#low_complexity (matched complement)

Seq primer: M13 FORWARD

POLYA=Yes.

FEATURES

SOURCE

location/Qualifiers

1..596

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-CF-DUI-abc-a-10-0-UI"

/tissue_type="Primary Lung Epithelial Cells"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_11b="UI-CF-DUI"

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a

modified polylinker; Site_1: BcoR I; Site_2: Not I;

UI-CF-DUI is a normalized cDNA library containing the

following tissue(s): Primary Lung Epithelial Cells The

library was constructed according to Bonaldo, Lennon and

Soares, Genome Research, 6:791-806, 1996. First strand

cDNA synthesis was primed with an oligo-dT primer

containing a Not I site. Double stranded cDNA was ligated

to an EcoR I adaptor, digested with Not I, and cloned

directionally into pT73-Pac vector. The oligonucleotide

used to prime the synthesis of first-strand cDNA contains

a library tag sequence that is located between the Not I

site and the (dT)18 tail. The sequence tag for this

library is GGCTGTAGGC.

TAG_L1b=UI-CF-DUI

TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368

TAG_SEQ=GGCTGTAGGC"

BASE COUNT 175 a 86 g 198 t

ORIGIN

Query Match

Best Local Similarity 12.9%; Score 129.2; DB 12; Length 596;

Matches 134; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY	860	CTTTCTTAAGGAAGTCACTGAGGCAACAGTGTAGAGAGAGAGAGAGACTCAA	919
DB	440	CTTACTTCACAACTACAGTACGTAGGCAACAGTGTAGAGAGAGAGAGACTCAA	381
QY	920	CCAACTAGAGCAGGATTAAGTTCCCTGTTCACTTTTCAGTGTCTGGAATGGA	979
DB	380	CCAACTAGAGCAGGATTAAGTTCCCTGTTCACTTTTCAGTGTCTGGAATGGA	321
QY	980	TGATTACCTCATTTGTCAAGTAG	1001
DB	320	TGATTACCTCATTTGTCAAGTAG	299

Search completed: December 16, 2003, 17:31:43
Job time : 2064.06 secs


```

Db      314 ATTAAGGAACTATTGAGGAGCTCCATATGATGCTGGAGAGCAAGCTACTTTTAAA 373
Qy      121 TTACTCTCGAATATGAGATGCAATTTGCGGATGATGATGAAAGCGCTCT 180
Db      374 TTACTCTCGAATATGAGATGCAATTTGCGGATGATGATGAAAGCGCTCT 433
Qy      181 GCTGTGCGCCGAGATTTCCACTTGAACCTTAAATGACTGATTCAGCTCTATGGAATT 240
Db      434 GCTGTGCGCCGAGATTTCCACTTGAACCTTAAATGACTGATTCAGCTCTATGGAATT 493
Qy      241 TATGTAACTTCTGGGAGAGGAAATATGCACTCCACAGATGCTTGTTCAGTTATGTC 300
Db      494 TATGTAACTTCTGGGAGAGGAAATATGCACTCCACAGATGCTTGTTCAGTTATGTC 553
Qy      301 TATGAGGCGCCCACTGCGAGATATGAGAGCCCACTCCCGGATATGAGAGCCCACTGCA 360
Db      554 TATGAGGCGCCCACTGCGAGATATGAGAGCCCACTCCCGGATATGAGAGCCCACTGCA 613
Qy      361 GGATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAGCCCACT 420
Db      614 GGATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAGCCCACT 673
Qy      421 GTGCGATATGAGAGCCCACTCTTGATATGAGAGCCCACTGCGAGATATGAGAGCCCA 480
Db      674 GTGCGATATGAGAGCCCACTCTTGATATGAGAGCCCACTGCGAGATATGAGAGCCCA 733
Qy      481 CCTCTAGAGATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGCGATATGAGAGCC 540
Db      734 CCTCTAGAGATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGCGATATGAGAGCC 793
Qy      541 CCACCTCTCGAATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAG 600
Db      794 CCACCTCTCGAATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAG 853
Qy      601 GCTCTACCTGCTGATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAG 660
Db      854 GCTCTACCTGCTGATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAG 913
Qy      661 AACGAGGCTTCTCTTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
Db      914 AACGAGGCTTCTCTTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 958

RESULT 3
BC022549      2266 bp      mRNA      linear      PRI 04-FEB-2002
LOCUS      Homo sapiens, Similar to RIKEN cDNA 4930521123 gene, clone
DEFINITION      MGC:26828 IMAGE:4815849, mRNA, complete cde.
ACCESSION      BC022549
VERSION      BC022549.1 GI:18490710
KEYWORDS      MGC.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 2266)
AUTHORS      Strausberg, R.
TITLES      Direct Submission
JOURNAL      Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK      NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT      Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Bromberg (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site:      http://www-shgc.stanford.edu

```

Contact: (Dickson, Mark) mcd@pax1.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov
 Series: IRAC Plate: 32 Row: m Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES

Source

Location/Qualifiers

1..2266

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:26828 IMAGE:4815849"

/issue_type="Brain, hippocampus"

/clone_id="NH_MGC_95"

/lab_host="DH10B"

/note="Vector: pBluescript"

32..634

/codon_start=1

/product="Similar to RIKEN cDNA 4930521123 gene"

/protein_id="AAH22549.1"

/db_xref="GI:18490711"

/translation="MAVNSHTNRGALIPNGESILKSPNYELSPORBSGNSVPSGRKTLFLTSYVIFITSCISIDPMSIPMPDLMNLTVPDPVPAAPTKGTLQAPYSGEQATPKIVFRNGALIERAQLMVTAAASAAARFPIRTLDNWSNGIYITRKGIQALHRCVLQVLSMEPHLDPMSEPHLPDTEPHLDMEPN"

BASE COUNT 709 a 477 c 468 g 612 t

ORIGIN

Query Match 97.4%; Score 686.6; DB 9; Length 2266;
 Best Local Similarity 99.3%; Pred. No. 5.7e-168;
 Matches 700; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

```

Qy      1 ATGCCATTGATCTGATGAGAGCACTCTGATGAGAGCACTGATGAGAGCACTGATGAGAGCACTG 60
Db      254 ATGCCATTGATCTGATGAGAGCACTCTGATGAGAGCACTGATGAGAGCACTGATGAGAGCACTG 313
Qy      61 ATTAAGGAACTATTGAGGAGCTCCATATGATGCTGGAGAGCAAGCTACTTTTAAA 120
Db      314 ATTAAGGAACTATTGAGGAGCTCCATATGATGCTGGAGAGCAAGCTACTTTTAAA 373
Qy      121 TTACTCTCGAATATGAGATGCAATTTGCGGATGATGATGAAAGCGCTCT 180
Db      374 TTACTCTCGAATATGAGATGCAATTTGCGGATGATGATGAAAGCGCTCT 433
Qy      181 GCTGTGCGCCGAGATTTCCACTTGAACCTTAAATGACTGATTCAGCTCTATGGAATT 240
Db      434 GCTGTGCGCCGAGATTTCCACTTGAACCTTAAATGACTGATTCAGCTCTATGGAATT 493
Qy      241 TATGTAACTTCTGGGAGAGGAAATATGCACTCCACAGATGCTTGTTCAGTTATGTC 300
Db      494 TATGTAACTTCTGGGAGAGGAAATATGCACTCCACAGATGCTTGTTCAGTTATGTC 552
Qy      301 TATGAGGCGCCCACTGCGAGATATGAGAGCCCACTCCCGGATATGAGAGCCCACTGCA 360
Db      553 TATGAGGCGCCCACTGCGAGATATGAGAGCCCACTCCCGGATATGAGAGCCCACTGCA 612
Qy      361 GGATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAGCCCACT 420
Db      613 GGATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAGCCCACT 672
Qy      421 GTGCGATATGAGAGCCCACTCTTGATATGAGAGCCCACTGCGAGATATGAGAGCCCA 480
Db      673 GTGCGATATGAGAGCCCACTCTTGATATGAGAGCCCACTGCGAGATATGAGAGCCCA 732
Qy      481 CCTCTAGAGATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGCGATATGAGAGCC 540
Db      733 CCTCTAGAGATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTGCGATATGAGAGCC 792
Qy      541 CCACCTCTCGAATATGAGAGCCCACTGCGAGATATGAGAGCCCACTGCGAGATATGAGAG 600

```



```

DB      793 CCACCTCTGAAATGAGCCCACTGACGAGAAATGAGCCCGCTCGGATACACA 852
QY      601 GCCTCACTGATGATCAGAGCCAGGCTCAGAAATTCACAGACGCCAGGCTCTGAA 660
DB      853 GCCTCACTGATGATCAGAGCCAGGCTCAGAAATTCACAGACGCCAGGCTCTGAA 912
QY      661 AACGAGCTTCTCTCTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
DB      913 AACGAGCTTCTCTCTCTGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 957

RESULT 4
HS250D10
LOCUS   220895 bp      DNA      linear      PRI 05-JUN-2003
DEFINITION Human DNA sequence from clone CTA-250D10 on chromosome 22 Contains
            the genes for SREBF2 (sterol regulatory element binding
            transcription factor 2), NAGA (alpha-N-acetylglucosaminidase), a
            gene similar to neuronal-specific septin 3, a pseudogene similar to
            ANT2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a
            genomic marker D2S1178, a CA repeat polymorphism, ESTs and a Cpg
            island, complete sequence.
ACCESSION 299716
VERSION 299716.4 GI:4456457
KEYWORDS HTG; ANT2; Cpg Island; D2S1178; NAGA; septin 3; SREBF2.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
            Bunkyoja; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 220895)
            Clark, G.
            Direct Submission
            Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
            On Mar 21, 1999 this sequence version replaced gi:1464339.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep CTA-250D10 is
            from the human BAC library described in U-J. Kim et al. (1996)
            Genomics 34, 213-218.
            VECTOR: pBAC108L
            ----- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquerry@sanger.ac.uk
            -----
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            This sequence is the entire insert of clone CTA-250D10 The true
            left end of clone RP1-18601 is at 129979 in this sequence. The true
            left end of clone RP3-359016 is at 1335 in this sequence. The true
            right end of clone RP5-821D11 is at 23458 in this sequence. The
            true right end of clone RP3-359016 is at 118711 in this sequence.

```

```

FEATURES
    source
        Location/Qualifiers
            1..220895
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="22"
                /clone="CTA-250D10"
                /clone_11b="CIT978SK-A1"
            17..29
                /note="2.6 copies 5 mer ATTG 26% conserved"
            33..325
                /note="AluX repeat: matches 1..301 of consensus"
            326..338
                /note="2.6 copies 5 mer TTGA 26% conserved"
            667..683
                /note="2.1 copies 8 mer AGAGACC 34% conserved"
            740..751
                /note="2.4 copies 5 mer TTTTA 24% conserved"
            1443..1457
                /note="7.5 copies 2 mer AC 21% conserved"
            1715..1997
                /note="AluY repeat: matches 5..293 of consensus"
            2103..2378
                /note="AluY repeat: matches 2..296 of consensus"
            2844..2855
                /note="12.0 copies 1 mer A 24% conserved"
            complement(2884..3033)
                /note="MIR repeat: matches 70..235 of consensus"
            complement(3040..3344)
                /note="AluY repeat: matches 1..310 of consensus"
            3446..3681
                /note="AluY repeat: matches 5256..5471 of consensus"
            3746..3837
                /note="AluY repeat: matches 5615..5707 of consensus"
            3838..4142
                /note="AluY repeat: matches 2..305 of consensus"
            4143..4441
                /note="AluY repeat: matches 5707..5999 of consensus"
            4442..4747
                /note="AluY repeat: matches 1..292 of consensus"
            4748..4858
                /note="AluY repeat: matches 5999..6110 of consensus"
            4881..5172
                /note="AluX repeat: matches 1..296 of consensus"
            5418..5429
                /note="2.4 copies 5 mer GCTGA 24% conserved"
            complement(5769..6021)
                /note="match: GSS: Em:A0069839"
            6074..6089
                /note="2.0 copies 8 mer CCCGCTG 32% conserved"
            6139..6153
                /note="2.5 copies 6 mer TTTTA 21% conserved"
            6141..6155
                /note="3.0 copies 5 mer TTTTA 21% conserved"
            complement(6366..6572)
                /note="MIR repeat: matches 5..214 of consensus"
            6839..6852
                /note="2.3 copies 6 mer GGAAT 28% conserved"
            7088..7114
                /note="5.8 copies 4 mer TTTA 36% conserved"
            7091..7109
                /note="1.9 copies 10 mer ATTATTA 38% conserved"
            complement(7122..7418)
                /note="AluX repeat: matches 1..300 of consensus"
            7464..7502
                /note="19.5 copies 2 mer AT 78% conserved"
            7502..7524
                /note="11.5 copies 2 mer AC 46% conserved"
            7523..7538
                /note="2.0 copies 8 mer CAAATAT 32% conserved"
            7535..7563
                /note="14.5 copies 2 mer AT 49% conserved"
            complement(7564..7859)

```



```

Oy 61 ATTAAGGAATATTACGACGCTCATATGTGGTGGGAAGACAAGACTACTTTAA 120
Db 318 ATTAAGGAACATTCACGACGCTCATATGTGGTGGGAAGACAAGACTACTTTAA 377
Oy 121 TTAGTCTTCAAAATGAGATGCGCATTTGATTTGCGGATGATGATGATGATGAT 180
Db 378 TTAGTCTTCAAAATGAGATGCGCATTTGATTTGCGGATGATGATGATGATGAT 437
Oy 181 GGTGTGTCGGGAGATTTCCACTTGAACCTTAATGATGATGATGATGATGATGAT 240
Db 438 GGTGTGTCGGGAGATTTCCACTTGAACCTTAATGATGATGATGATGATGATGAT 497
Oy 241 TATGTAATTAATGAGGGAAT---ATGTCACCTCCACAGATGCTTTGAG----- 292
Db 498 TACATTAATTAATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 557
Oy 293 ---TTATTTGCTATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 348
Db 558 CCAATTTGATTAATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 617
Oy 349 GCGCCACCTGAGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 408
Db 618 ACTCCACCAAGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 677
Oy 409 AGAGCTGACCTGTGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 468
Db 678 GAGGCTGCTGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 737
Oy 469 TATGAGGCTGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 528
Db 738 TATGAGGCTGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 797
Oy 529 GATATGAGGCTGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 588
Db 798 GATATGAGGCTGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 857
Oy 589 GGGGATATGAGGCTGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 648
Db 858 GGTGAAATGAGGCTGATATGAGGGAATGAGTGTCTCTCCACAGATGCTTTGAG----- 917
Oy 649 -----CAGGCTCTGAAAGAGGCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 696
Db 918 TCTCAGATCTATGAGGCTGATATGAGGGAATGAGTGTCTCTCTCTCTCTCTCT 977
Oy 697 GTCCATT 703
Db 978 GTCCATT 984

```

```

RESULT 6
BX296515 179222 bp DNA 1linear HTG 26-MAY-2003
LOCUS
DEFINITION
unorderd pieces.
ACCESSION
BX296515.6 GI:31076160
VERSION
HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Sus scrofa (pig)
SOURCE
Sus scrofa
ORGANISM
Sus scrofa
Mammalia; Eutheria; Cetartiodactyla; Suina, Suidae; Sus.
1 (bases 1 to 179222)
Tracey, A.
Direct Submission
Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On May 26, 2003 this sequence version replaced gi:31043704.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk

```

```

Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: br121d21
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 177502 bases at least Q40
Consensus quality: 177971 bases at least Q30
Consensus quality: 178283 bases at least Q20
Insert size: 167315; 12.1% error; agarose-tp
Quality coverage: 7,82x in Q20 bases; sum-of-contigs quality
coverage: 8.59x in Q20 bases; agarose-tp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5850: contig of 5850 bp in length
* 5851 5950: gap of 100 bp
* 5951 74264: contig of 68314 bp in length
* 74265 74365: gap of 100 bp
* 74365 84656: contig of 10292 bp in length
* 84657 84757: gap of 100 bp
* 84757 108493: contig of 23737 bp in length
* 108494 108593: gap of 100 bp
* 108594 114454: contig of 5661 bp in length
* 114455 114554: gap of 100 bp
* 114555 179222: contig of 64668 bp in length.
FEATURES
source
1..179222
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/clone="PigR-121D21"
/clone_11b="PigR"
1..5850
/note="assembly fragment:02838
fragment_chain:1"
5951..74264
/note="assembly fragment:01479
fragment_chain:1"
74365..84656
/note="assembly fragment:01113
fragment_chain:1"
84757..108493
/note="assembly fragment:00846
fragment_chain:1"
108594..114454
/note="assembly fragment:01058
fragment_chain:2"
114555..179222
/note="assembly fragment:02416
fragment_chain:2"
BASE COUNT 44892 a 44880 c 44117 g 44831 t 502 others
ORIGIN
Query Match 30.5% Score 215; DB 2; Length 179222;
Best Local Similarity 73.0%; Pred. No. 6.9e-45;
Matches 294; Conservative 0; Mismatches 100; Indels 9; Gaps 1;
Oy 301 TATGAGGCTGATATGAGGGAATGAGTGTCTCTCTCTCTCTCTCTCTCTCTCT 360
Db 55424 TACGAGGCTGATATGAGGGAATGAGTGTCTCTCTCTCTCTCTCTCTCTCTCT 55483
Oy 361 GATATGAGGCTGATATGAGGGAATGAGTGTCTCTCTCTCTCTCTCTCTCTCTCT 420
Db 55484 GATATGAGGCTGATATGAGGGAATGAGTGTCTCTCTCTCTCTCTCTCTCTCTCT 55543

```



```

* 172055 172154: gap of 100 bp
* 172155 175210: contig of 3056 bp in length
* 175211 175310: gap of 100 bp
* 175311 179379: contig of 4069 bp in length
* 179380 179479: gap of 100 bp
* 179480 182698: contig of 3219 bp in length
* 182699 182798: gap of 100 bp
* 182799 186779: contig of 3981 bp in length
* 186780 186879: gap of 100 bp
* 186880 195626: contig of 8747 bp in length
* 195627 195726: gap of 100 bp
* 195727 199396: contig of 3670 bp in length
* 199397 199496: gap of 100 bp
* 199497 206412: contig of 6916 bp in length
* 206413 206512: gap of 100 bp
* 206513 209619: contig of 3107 bp in length
* 209620 209719: gap of 100 bp
* 209720 219083: contig of 9364 bp in length
* 219084 219184: gap of 100 bp
* 219184 222469: contig of 3286 bp in length.

FEATURES
    source             Location/Qualifiers
        1..222469      /organism="Danio rerio"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7955"
                        /clone_1fb="Dantiokey"
        1..6656         /note="assembly_fragment:01359
                        fragment_chain:1"
        6757..10255     /note="assembly_fragment:01385
                        fragment_chain:1"
        10356..13223    /note="assembly_fragment:00592
                        fragment_chain:1"
        13324..19935    /note="assembly_fragment:00237
                        fragment_chain:1"
        20036..32597    /note="assembly_fragment:00949
                        fragment_chain:1"
        22698..38569    /note="assembly_fragment:00615
                        fragment_chain:1"
        28670..37965    /note="assembly_fragment:01306
                        fragment_chain:2"
        38066..41140    /note="assembly_fragment:00487
                        fragment_chain:2"
        41241..45108    /note="assembly_fragment:00232
                        fragment_chain:2"
        45209..47730    /note="assembly_fragment:00935
                        fragment_chain:2"
        47831..51884    /note="assembly_fragment:01415
                        fragment_chain:2"
        51985..55857    /note="assembly_fragment:00878
                        fragment_chain:3"
        55958..60498    /note="assembly_fragment:01208
                        fragment_chain:3"
        60599..63641    /note="assembly_fragment:01266
                        fragment_chain:3"
        63742..72674    /note="assembly_fragment:01329
                        fragment_chain:3"
        72775..75602

```

```

/note="assembly_fragment:01183
fragment_chain:4"
75703..78062
misc_feature /note="assembly_fragment:01179

```

```

Query Match 30.5% Score 215; DB 2; Length 222469;
Best Local Similarity 73.0%; Pred. No. 7,1e-45;
Matches 294; Conservative 0; Mismatches 100; Indels 9; Gaps 1;

```

```

QY 301 TATGGGCGCCCACTGCGAGATATGAGAGCCCACTCCGATACGAGCCCACTGCA 360
DB 197080 TACGAGGCCCACTGAAAGACTGAAAGCCCACTGAGATATGAGCCCACTGCG 197139
QY 361 GATATGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGCA 420
DB 197140 GATATGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGCA 197199
QY 421 GTGCGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCA 480
DB 197200 GAAAGACAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCA 197259
QY 481 CCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCC 540
DB 197260 CCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCC 197319
QY 541 CCACCTCGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCC 600
DB 197320 CCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCC 197379
QY 601 GCCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCC 660
DB 197380 GCCCATGCTGCTGAGAAATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAGCC 197432
QY 661 AACGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 703
DB 197433 --GGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 197473

```

```

RESULT 8
BX296540/c 129624 bp DNA linear HTG 02-APR-2003
LOCUS Sus scrofa clone PlgR-231X18, *** SEQUENCING IN PROGRESS ***, 37
DEFINITION
ACCESSION BX296540
VERSION BX296540.2 GI:29500961
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 129624)
Burton,J.
Direct Submission
Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquist@
humgeny@sanger.ac.uk
On Apr 2, 2003 this sequence version replaced gi:29353441.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humgeny@sanger.ac.uk
----- Project Information
Center project name: b7231X18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 113740 bases at least Q40
Consensus quality: 118187 bases at least Q30
Consensus quality: 120712 bases at least Q20
Insert size: 126024; sum-of-contigs
Insert size: 154528; 4.6% error; aggrose-fp
Quality coverage: 2.35x in Q20 bases; sum-of-contigs Quality

```



```

misc_feature      fragment chain:7"
                    62661..64963
                    /note="assembly_fragment:00016"
misc_feature      65064..67191
                    /note="assembly_fragment:00065"
misc_feature      67292..72005
                    /note="assembly_fragment:00070"
misc_feature      72106..76595
                    /note="assembly_fragment:00136"
misc_feature      76696..79825
                    /note="assembly_fragment:00162"
misc_feature      79926..83540
                    /note="assembly_fragment:00367"
misc_feature      83641..86276
                    /note="assembly_fragment:00492"
misc_feature      86377..94620
                    /note="assembly_fragment:00545"
misc_feature      94721..97520
                    /note="assembly_fragment:00745"
misc_feature      97621..99784
                    /note="assembly_fragment:00801"

```

Query Match 29.6%; Score 208.4; DB 2; Length 129624;

Best Local Similarity 74.3%; Pred. No. 3.5e-43;

```

Matches 263; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

OY 296 TTGCTATGGGCCCCCACCCTGAGATATGAGCCCCCAGATACGAGCCCAAC 355
DB 65475 TAGGATATGAGAGGCCAACAGAGATATAGAGCCCACTTCAGATACGAGCCCAAC 65416
OY 356 CTGCAGATATGAGAGCCCAACCCGATGAGAAAGAGCCCGCTGGATACAGAGCT 415
DB 65415 CTCCAGATATGAGAGCCCACTTCAGATATGAGAGCCCACTTCAGATATGAGAGCT 65356
OY 416 CACCTGTGATATGAGAGCCCACTTCGATATGAGAGCCCACTTCAGATATGAGAG 475
DB 65355 CACCTGTGATATGAGAGCCCACTTCGATATGAGAGCCCACTTCAGATATGAGAG 65296
OY 476 CCCCACCTCTAGATATGAGAGCCCACTTCGATATGAGAGCCCACTTCGATATG 535
DB 65295 CCCCACCTCTAGATATGAGAGCCCACTTCGATATGAGAGCCCACTTCGATATG 65236
OY 536 GAGCCCACTCTGATATGAGAGCCCACTTCGATATGAGAGCCCACTTCGATATG 595
DB 65235 GAGCCCACTCTGATATGAGAGCCCACTTCGATATGAGAGCCCACTTCGATATG 65176
OY 596 ACAGAGCCCTCAGCTGATCAGAGCCCACTTCAGATATGAGAGCCCACTTCAGAG 649
DB 65175 ATGAGAGCCCACTCAGATATGAGAGCCCACTTCAGATATGAGAGCCCACTCAG 65122

```

RESULT 9
AC113593 168425 bp DNA linear HTG 06-JUN-2002
LOCUS Mus musculus clone RP23-363124, WORKING DRAFT SEQUENCE, 18 ordered
DEFINITION pieces.

AC113593 GI:21327432
VERSION HTG: HTGS PHASE2: HTGS DRAFT: HTGS FULLTOP.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 168425)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-363124
JOURNAL Unpublished
2 (bases 1 to 168425)

REFERENCE 2 (bases 1 to 168425)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, A., Chang, J., Chazaro, B.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,

TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (03-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 168425)

REFERENCE
AUTHORS
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:21313869.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: 123816
Center clone name: 3631_24
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 163227 bases at least Q40
Consensus quality: 165972 bases at least Q30
Insert size: 16200; agarose-1p
Insert size: 166725; sum-of-coverage
Quality coverage: 8.6 in Q20 bases; agarose-1p
Quality coverage: 8.3 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently
consists of 18 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces

```

* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
* 1413 1512: gap of 100 bp in length
* 1513 1512: gap of 100 bp in length
* 3367 3366: contig of 1854 bp in length
* 3467 3466: gap of 100 bp in length
* 5543 5542: contig of 2076 bp in length
* 5643 5642: gap of 100 bp in length
* 7950 7949: contig of 2307 bp in length
* 8050 8049: gap of 100 bp in length
* 42610 42610: contig of 34561 bp in length
* 42611 42710: gap of 100 bp in length
* 42711 42717: contig of 3007 bp in length
* 45718 45817: gap of 100 bp in length
* 45818 50106: contig of 4289 bp in length
* 50206 50206: gap of 100 bp in length
* 52907 52906: contig of 2739 bp in length
* 52946 53045: gap of 100 bp in length
* 53046 56269: contig of 3224 bp in length
* 56270 56369: gap of 100 bp in length
* 56370 59897: contig of 3528 bp in length
* 59898 59897: gap of 100 bp in length
* 59998 67207: contig of 7210 bp in length
* 67208 67307: gap of 100 bp in length
* 67308 78743: contig of 11436 bp in length
* 78744 78843: gap of 100 bp in length
* 78844 89991: contig of 11148 bp in length
* 89992 90091: gap of 100 bp in length
* 90092 107918: contig of 17827 bp in length
* 107919 108018: gap of 100 bp in length
* 108019 124874: contig of 16856 bp in length
* 124875 124974: gap of 100 bp in length
* 124975 142093: contig of 17119 bp in length
* 142094 142193: gap of 100 bp in length
* 142194 167111: contig of 24918 bp in length
* 167112 167211: gap of 100 bp in length
* 167212 168425: contig of 1214 bp in length.
Location/Qualifiers
1. 168425
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone.lib="RPC1-23 Female Mouse BAC"
1. 1412
/note="assembly_fragment"
misc_feature
1513. 3366
/note="assembly_fragment"
misc_feature
3467. 5542
/note="assembly_fragment"
misc_feature
5643. 7949
/note="assembly_fragment"
misc_feature
8050. 42610
/note="assembly_fragment"
misc_feature
42711. 45717
/note="assembly_fragment"
misc_feature
45818. 50106
/note="assembly_fragment"
misc_feature
50207. 52945
/note="assembly_fragment"
misc_feature
53046. 56269
/note="assembly_fragment"
misc_feature
56370. 59897
/note="assembly_fragment"
misc_feature
59998. 67207
/note="assembly_fragment"
misc_feature
67308. 78743
/note="assembly_fragment"
misc_feature
78844. 89991

```

```

/note="assembly_fragment"
90092. 107918
/note="assembly_fragment"
108019. 124874
misc_feature
108019. 124874
/note="assembly_fragment"
124975. 142093
misc_feature
142194. 167111
/note="assembly_fragment"
167212. 168425
misc_feature
167212. 168425
/note="assembly_fragment"
168425. 1700
vector side:right"
BAC COUNT 44391 a 39572 c 38747 g 44015 t 1700 others
ORIGIN
Query Match 25.4%; Score 179; DB 2; Length 168425;
Best Local Similarity 65.3%; Pred. No. 1.6e-35;
Matches 263; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 303 TGGGGCCCCCACTGCAAGATATGAGAGCCCACTCCCGATGAGAGCCCACTGCAAG 362
DB 66293 TGGATACGAGCTGTGAGATATGAGCTCCCTCTCTATATGATACCTACCAATGAG 66352
QY 363 ATATGAGCCCAACCCGAGAAATGAAAGCCCGCTGTGAGATACAGGCTCACTGT 422
DB 66353 CTATGAGATTCACCTCTCTGTATATGAGCTCCACCTGTATATGAGATACCACTCC 66412
QY 423 GCGATATGAGCCCACTCTGTATATGAGAGCCCACTGCAAGATATGAGAGCCCACT 482
DB 66413 TGGATATGAGAGCCCACTATATGAGATATGAGAGCCCACTGCAAGATATGAGAGC 66472
QY 483 TCTGATATGAGAGCCCACTCTGTATATGAGAGCCCACTGCAAGATATGAGAGCC 542
DB 66473 TATGAGATCCCGTCTCCCACTCCCAATATGAGAGCCCGCTATGAGATGAGACTCC 66532
QY 543 ACCCTGAGATATGAGAGCCCACTGCAAGAAATGAAAGCCCGCTGTGAGATACAGC 602
DB 66533 TCTTCAAGAGCGTATGATATCTCTGTATCAAGAGCCCACTGTATGAGAGAGC 66592
QY 603 CTCACCTGATATGAGAGCCCACTGCAAGATATGAGAGCCCACTGCAAGAGC 662
DB 66593 TCACCTGCTGTGATATGAGAGCCCACTGCAAGATATGAGAGC 66552
QY 663 CGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
DB 66653 CCAGGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 66695
RESULT 10
AC104325 224086 bp DNA linear ROD 30-MAY-2003
LOCUS AC104325
DEFINITION Mus musculus clone rp23-204m3 map 15 strain C57BL/6J, complete
sequence.
ACCESSION AC104325
VERSION AC104325.28 GI:31193955
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 224086)
AUTHORS Jiang,H., Song,L. and Roe,B.A.
TITLE Mus musculus BAC Clone rp23-204m3
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 224086)
REFERENCE Jiang,H., Song,L. and Roe,B.A.
TITLE Direct Submission
AUTHORS Submitted (10-DEC-2001) Department Of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 224086)
AUTHORS Jiang,H., Song,L. and Roe,B.A.

```


TITLE Direct Submission
JOURNAL Submitted (19-MAY-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 224086)
AUTHORS Jiang, H., Song, L. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 5 (bases 1 to 224086)
AUTHORS Jiang, H., Song, L. and Roe, B. A.
TITLE Direct Submission
JOURNAL Submitted (30-MAY-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT ----- Genome Center
Center: Department of Chemistry And Biochemistry
The University Of Oklahoma
Center code: UOKNOR

FEATURES
source 1. 224086
Location/Qualifiers
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="15"
/clone="TP23-204m3"
/clone_1lb="RPCT - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT 55592 a 55124 c 56095 g 57275 t
ORIGIN
Query Match 25.4%; Score 179; DB 10; Length 224086;
Best Local Similarity 65.3%; Pred. No. 1.7e-35;
Matches 263; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 303 TGGGGCCCACTGAGATATGAGAGCCCACTCCCGATACGAGCCCACTGAGAG 362
DB 162839 TGGATACGAGAGCTGATGATATGAGTCCCTCTCTCTATATGATGATACCCCAATGAG 162898
QY 363 ATATGAGGCCCAACCGTATGAGAAATGAGAGCCCGCTTGGATACAGAGCTTCACTGT 422
DB 162899 CTATGAGAGTTCACCTCTCGATATGAGACCTTCACTGATGAGATGAGATCCCACTCC 162958
QY 423 GGGATATGAGAGCCCACTCTTGGATATGAGAGCCCACTGAGATGAGAGCCCACTCC 482
DB 162959 TGGATATGAGAGCCCACTATGAGATGAGAGCCCACTCCAGATACGAGATCTACAC 163018
QY 483 TCTAGATATGAGAGCCCACTCTTGGATATGAGAGCCCACTCTCGATATGAGAGCCCA 542
DB 163019 TATGAGATCCGATTCCTCCCACTCCAGATATGAGAGCCCGCTATGAGGATGAGACTCC 163078
QY 543 ACCTCTCGATATGAGAGCCCACTCTGAGAAATGAGAGCCCGCTTGGATACAGAGC 602
DB 163079 TCTTTCAGAGGCGGATGATCTACCTCTGATCAAGGCGCACTCTGTAGCAGAGAAAGC 163138
QY 603 CTCACCTGCTGATACGAGAGCCGAGGCTCAGATCTACAGAGCCCAAGGCTCTGAAAA 662
DB 163139 TCACCTGCTGAGGATGAGAGAGCCCACTATGCTGAGGAGCTCAGAGATCTGAAATT 163198
QY 663 CGAGGCTTCTCTCCCTGCT 705
DB 163199 CGAGGATCTTCT 163241

RESULT 11
AC107527 253149 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-92M24, WORKING DRAFT SEQUENCE, 3
DEFINITION unoriented pieces.
ACCESSION AC107527

VERSION
KEYWORDS AC107527.5 GI:30580771
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Buxarjota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 253149)
AUTHORS Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alldbrooks, S., Amin, A., Anguiano, D.,
Aryalchek, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Bennahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, T., Caesar, H., Center, A.,
Cheval, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M. L., Davis, C., Davy-Carroll, L., De Andrade, C., Dederich, D.,
Delgado, O., Denson, S., Derramo, C., Ding, Y., Dinh, H., Divya, K.,
Drepper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duval, B., Baves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kovits, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshew, L., Louised, H., Lozada, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, R.,
Manning, S., McLeod, M. P., McNeill, T. Z., Meenen, B.,
Milosavljevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munsada, M., Murphy, M., Nat, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
Pascernek, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poldexter, A., Popovic, D., Primus, B., Pu, L.,
Puro, M., Qutroz, J., Rachlin, R., Reeves, K., Reiser, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivas, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shattman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Slisom, I., Sitter, C. D., Smaj, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svetek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,
Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,
Williams, G., Wilson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yahub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.

TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 253149)
AUTHORS Worley, K. C.
TITLE Direct Submission
JOURNAL Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 253149)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:23664661.
The sequence in this assembly is a combination of BAC based reads

Plopper, F., Poindexter, A., Popovic, D., Primus, B., Pu, L.-L., Puzos, M., Quiroz, J., Rachlin, R., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajic, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Thor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejo, Z., Umami, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wleczek, R., Woodson, H., Morley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Unpublished
2 (bases 1 to 270171)
Rat Genome Sequencing Consortium:
Direct Submission
Submitted (05-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 270171)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced g123908490.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with 'Ns' to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBOD
Center clone name: CH230-327L20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 251525 bases at least Q40
Consensus quality: 254049 bases at least Q30
Consensus quality: 255599 bases at least Q20
Estimated insert size: 259926; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 270171: contig of 270171 bp in length.
Location/Qualifiers

source 1. 270171
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-327L20"

misc_feature
1. 2686
/note="wgs end extension
clone end:5p5"

misc_feature
4191..5073
/note="clone boundary
clone end:5p6
site:
end, sequence: B2160678"
complement (267941..268827)
/note="clone boundary
clone end:17
site:
end, sequence: B2160665"

misc_feature
/note="clone boundary
clone end:5p6
site:
end, sequence: B2160678"
complement (267941..268827)
/note="clone boundary
clone end:17
site:
end, sequence: B2160665"

BASE COUNT 67733 a 63574 c 62822 g 63249 t 12793 others
ORIGIN

Query Match 24.8%; Score 174.6; DB 2; Length 270171;
Best Local Similarity 64.4%; Pred. No. 2.4e-34;
Matches 261; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 301 TATGAGGCCCCACCTGAGAGATATGAGCCCACTCCGAGATGAGAGCCCACTGCA 360
DB 15227 TCTGATATCGAGAGCTGAGATATGATCCCTCTCTCTATATGATGATCCCTATG 15168
QY 361 GGATATGAGGCCCAACCCGTAGAAATGAAAGCCCGCTGAGATACAGAGCTCACT 420
DB 15167 GGCATATGAGATCCACCTCCGAGATATGAGACCCCACTGAGATATGAGATCCCACT 15108
QY 421 GTGCGATATGAGGCCCACTCTTGAATPAGAGACCCCACTGAGATATGAGACCCA 480
DB 15107 CCTGATATGAGAGCCCACTATGACATATGAGCCCACTCTTACATATGATATCA 15048
QY 481 CCTGATATGAGAGCCCACTCTTGAATPAGAGACCCCACTCTGATATGAGAGCC 540
DB 15047 CCTATGATATCGATCTCCCACTCCGATATGAGACCCCACTATGATGATGATGAAAC 14988
QY 541 CCACCTCTGAGATATGAGGCCCACTGAGAAATGAAAGCCCGCTGAGAGATATGAG 600
DB 14987 CCGGCTCAGGAGATGATGATCATCTCTTGAATATCAATCCCTGCTGACAGAA 14928
QY 601 GCTTACCTGCTGATATGAGAGCCAGGCTCAGAAATCTACAGAGCCCACTCTGAA 660
DB 14927 GCACTTACCTGCTGATGATGAGAGCCAGGCTCAGAAATCTGAGAGCCCACTGAA 14868
QY 661 AAGCAGCTTCTCTTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
DB 14867 TTTCAGGATCTTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14823

RESULT 13
AC020698 148418 bp DNA linear PRI 01-MAR-2002
DEFINITION Homo sapiens BAC clone Kp11-45f23 from 4, complete sequence.
ACCESSION AC020698
VERSION AC020698.4 GI:11120934
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukharova, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 148418)
AUTHORS Sulston, J.B. and Waterston, R.
Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 148418)
AUTHORS Nguyen, C., Drone, K., Hawkins, M. and Ureta, M.

FEATURES

TITLE The sequence of Homo sapiens BAC clone RP11-45F23
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 148418)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 148418)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 5 (bases 1 to 148418)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 148418)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 148418)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 8, 2000 this sequence version replaced gl.7630812.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics -----
Center project name: H_NH0045F23

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, R.,
Tateno, M., Cataneese, J.J., and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pterer de Jong
and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-46G22. Actual start of
this clone is at base position 1 of RP11-45F23; actual end is at
base position 148418 of RP11-45F23.
Location/Qualifiers
I .148418

```
/organism="Homo sapiens"  
/mol_type="genomic DNA"  
/db_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-45F23"  
/clone_1kb="RPCI-11"  
1..434  
repeat_region  
/rpt_family="L1"  
507..807  
repeat_region  
/rpt_family="L1"  
807..2585  
repeat_region  
/rpt_family="L1"  
2616..3020  
repeat_region  
/rpt_family="L1"  
3062..3250  
repeat_region  
/rpt_family="L1"  
3251..3596  
repeat_region  
/rpt_family="MER2_type"  
3597..3843  
repeat_region  
/rpt_family="L1"  
9001..9137  
repeat_region  
/rpt_family="Alu"  
11353..11406  
repeat_region  
/rpt_family="ERV1"  
17191..17869  
repeat_region  
/rpt_family="L1"  
19869..20177  
repeat_region  
/rpt_family="Alu"  
21287..22345  
repeat_region  
/rpt_family="L1"  
22727..22937  
repeat_region  
/rpt_family="MER2_type"  
22933..23133  
repeat_region  
/rpt_family="MER2_type"  
23355..23442  
repeat_region  
/rpt_family="MIR"  
23485..23625  
repeat_region  
/rpt_family="MER2_type"  
24107..24872  
repeat_region  
/rpt_family="L1"  
24873..25320  
repeat_region  
/rpt_family="ERV1"  
25372..25522  
misc_feature  
/note="similar to EST BC087288 (NTD:98477683)"  
28473..28614  
repeat_region  
/rpt_family="ERV1"  
28622..28736  
repeat_region  
/rpt_family="ERV1"  
28790..28843  
repeat_region  
/rpt_family="ERV1"  
28837..29212  
repeat_region  
/rpt_family="ERV1"  
29215..29821  
repeat_region  
/rpt_family="ERV1"  
29823..29972  
repeat_region  
/rpt_family="ERV1"  
29974..30286  
repeat_region  
/rpt_family="ERV1"  
30287..30737  
repeat_region  
/rpt_family="ERV1"  
30745..31216  
repeat_region  
/rpt_family="L1"  
31227..31275  
repeat_region  
/rpt_family="MIR"  
31556..31612  
repeat_region  
/rpt_family="L2"  
32104..32229  
repeat_region  
/rpt_family="Alu"  
32252..32356  
repeat_region  
/rpt_family="L2"  
33845..33935  
repeat_region  
/rpt_family="MER81"
```

```

repeat_region 34001..34412
misc_feature 34405..34797
misc_feature 34408..35071
misc_feature 34408..34963
misc_feature 34408..34997
misc_feature 34408..34874
misc_feature 34408..34885
misc_feature 34408..34894
misc_feature 34408..34664
misc_feature 34408..34592
misc_feature 34411..34921
misc_feature 34412..34520
misc_feature 34413..34664
misc_feature 34414..34495
misc_feature 34415..35417
misc_feature 34416..34684
misc_feature 34417..34921
misc_feature 34418..34885
misc_feature 34423..34811

```

```

Query Match 18.9%; Score 133.4; DB 9; Length 148418;
Best Local Similarity 61.8%; Pred. No. 1.2e-23;
Matches 212; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```

```

QY 307 GCGCCACCTGAGATGAGAGCCGACCTCCGAGATGAGAGCCGACCTGAGATGAT 366
DB 43061 GCTTACCTCCAGAGATGCTTACCTCCAGAGATGCTTACCTCCAGAGATGC 43120
QY 367 GGAGCCCAACCCCTAGAGATGAGAGCCGCTGTGAGATGAGAGCCCTGTGCGA 426
DB 43121 TTAGCTCCAGCCGAGGCTTACCTCCAGAGATGCTTACCTCCAGAGATGC 43180
QY 427 TATGAGCCCACTCTTGTGATGAGAGCCGCTGTGAGATGAGAGCCGCTGTGCGA 486
DB 43181 TGCTTACCTCCAGAGATGCTTACCTCCAGAGATGCTTACCTCCAGAGATGC 43240
QY 487 GAGTATGAGAGCCGCTTGTGATGAGAGCCGCTGTGAGATGAGAGCCGCTGTGCGA 546
DB 43241 GAGTATGAGAGCCGCTTGTGATGAGAGCCGCTGTGAGATGAGAGCCGCTGTGCGA 43300
QY 547 CTGCGATGAGAGCCGCTTGTGATGAGAGCCGCTGTGAGATGAGAGCCGCTGTGCGA 606
DB 43301 CAGAGCTGCTTACCTCCAGAGATGCTTACCTCCAGAGATGCTTACCTCCAGAGATGC 43360
QY 607 CCGTGTGATGAGAGCCGCTTGTGATGAGAGCCGCTGTGAGATGAGAGCCGCTGTGCGA 649
DB 43361 CCGGAGATGCTTACCTCCAGAGATGCTTACCTCCAGAGATGCTTACCTCCAGAGATGC 43403

```

```

RESULT 14
AC103031/c 235381 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-203L17, WORKING DRAFT SEQUENCE, 3
DEFINITION unordered pieces.
ACCESSION AC103031

```

VERSION KEYWORDS SOURCE ORGANISM

AC103031.5 GI:30580672
HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE AUTHORS

1 (bases 1 to 235381)
Muzny D, Marle M, Metzker M, Lee A, Abramson S, Adams C, Alder J, Allen C, Allen H, Alebrooks S, Amin A, Angiano D, Ayalebech V, Ayagi A, Ayodeji M, Bacc B, Baden H, Baldwin D, Bandaranaike D, Barber M, Barnes M, Benhammed F, Biwalto K, Blair J, Blankensbury K, Blyth P, Brown M, Bryant N, Buhay C, Burch P, Burrell K, Calderon S, Cardenas V, Carter K, Cavazos I, Cessari H, Centes A, Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu J, Cleveland C, Cockrell R, Cox C, Coyle M, Cree A, D'Souza L, Davila M, L., Davis C, Davy-Carroll L, De Anda C, Dederich D, Delgado O, Denison S, Deramo C, Ding Y, Dinh H, Divya K, Draper H, Dugan-Rocha S, Dunn A, Durbin K, Duval B, Evans K, Egan A, Escotto M, Eugene C, Evans C, A., Falls T, Fan G, Fernandez S, Finley M, Flagg N, Forbes L, Foster M, Foster P, Fraser C, Gabisi A, Gant R, Garcia A, Garner T, Garza M, Gebregeorgis E, Geer K, Gill R, Grady M, Guerra M, Guevara M, Gunaratne P, Haaland W, Hamill C, Hamilton C, Hamilton K, Harvey Y, Havlak P, Hawes A, Henderson N, Hernandez J, Hernandez R, Hines S, Hladun S, L., Hodgson A, Hognes M, Hollins B, Howells S, Huliyil S, Hume J, Idlebird D, Jackson A, Jackson L, Jacob L, Jiang H, Johnson B, Johnson R, Jolivet A, Kapachy S, Kelly S, Khan Z, King L, Koyar C, Kovis C, Kraft C, L., Lebow H, Levan J, Lewis L, Li Z, Liu J, Liu J, Liu W, Liu Y, London P, Longacre S, Lopez J, Lorenshewa L, Louisedge H, Lozada R, J., Lu X, Ma J, Maheshwari M, Mahindartne M, Mahmoud M, Malloy K, Mangum A, Mangum B, Mapue P, Martin K, Martin R, Martinez E, Manwiny S, McLeod M, P., McNeill T, Z., Meenen E, Milosavljevic A, Miner G, Minja E, Montemayor J, Moore S, Morgan M, Morris K, Morris S, Mundasa M, Murphy M, Natir L, Nankervis C, Neal D, Newton N, Nguyen N, Norris S, Nwokilemeh O, Okunom G, Olarnunagoon A, Pal S, Parke K, Pasernak S, Paul H, Perez A, Perez L, Pfankoch C, Plopper F, Polidexter A, Popovic D, Frims R, Pu L, L., Puazo M, Quiroz J, Rachlin S, Reeves K, Regier M, A., Reigh R, Reilly B, Reilly M, Ren Y, Reuter M, Richards S, Riggs F, Rives C, Rodkey T, Rojas A, Rose M, Rose R, Ruiz S, J., Sanders M, Saverly G, Scherer S, Scott G, Shatman S, Shen H, Shetty J, Shvartsbeyn A, Sison I, Sitter C, D., Smajz D, Sneed A, Sodergren E, Song X, Z., Sorelle R, Soza J, Steimle M, Strong R, Sutton A, Svatek A, Taber P, Taylor C, Taylor T, Thomas N, Thomas S, Tingey A, Trejos Z, Uemami K, Valas R, Vera V, Villalana D, Waldron L, Walker B, Wang J, Wang Q, Wang S, Warren J, Warren R, Wei X, White F, Williams G, Wilson R, Wiczyski R, Woodem H, Worley K, Wright D, Wright R, Wu J, Yakub S, Yen J, Yoon L, Yoon V, Yu F, Zhang J, Zhou J, Zhou X, Zhao S, Dunn D, von Niederhausern A, Weiss R, Smith D, R., Holt R, A., Smith H, O., Weinerstock G, and Gibbs R. A.

TITLE JOURNAL AUTHORS TITL JOURNAL

Unpublished
2 (bases 1 to 235381)
Worley K. C.
Direct Submission
Submitted (13-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 235381)
Rat Genome Sequencing Consortium.

REFERENCE AUTHORS TITL JOURNAL

Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23123692.
The sequence in this assembly is a combination of BAC based reads

COMMENT

and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: G1ZG
Center clone name: CH230-203L17

----- Summary Statistics

Assembly program: Atlas 3.0;
Consensus quality: 221482 bases at least Q40
Consensus quality: 223612 bases at least Q30
Consensus quality: 224834 bases at least Q20
Estimated insert size: 231669; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 232989: contig of 232989 bp in length
* 232990 233089: gap of unknown length
* 233090 234257: contig of 1168 bp in length
* 234258 234357: gap of unknown length
* 234358 235381: contig of 1024 bp in length.

FEATURES

source

1. 235381
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-203L17"

misc_feature

1. 1102
/note="wgs end-extension
clone_end:Sp6"

misc_feature

4245. 4594
/note="clone boundary
clone_end:Sp6"

misc_feature

end_sequence: BH364165"
complement(230331..230677)
/note="clone boundary
clone_end:T7"

misc_feature

end_sequence: BH364164"
231434..232989
/note="wgs end-extension
clone_end:T7"

BASE COUNT 59674 a 53524 c 55897 g 56470 t 9816 others
ORIGIN

Query Match

Best Local Similarity 58.7%; Pval. No. 2.7e-19;
Matches 202; Conservative 0; Mismatch 142; Indels 0; Gaps 0;

302 ATGGAGCCCACTGATGATGAGCCCACTCCCGATACGAGCCCACTGACAG 361

Db 141772 ATGAGACCTCTCTATTACATGAGACCTCTTATTACATGAGACCTTCTAT 141713
Cy 362 GATATGAGGCCCAACCCGTGAAATGAGAGCCCGCTGAGATACAGCCCTACCTG 421
Db 141712 TACATGAGACCTTCTATTACATGAGACCTCTTATTACATGAGACCTTCTC 141653
Cy 422 TGGATATGAGGCCCAACCTCTGATATGAGAGCCCACTGAGATATGAGAGCCAC 481
Db 141652 TATTACATGAGACCTCTCTATTACATGAGATCAGCTCTTATTACATGAGACCTT 141593
Cy 482 CTCGAGATATGAGGCCCACTCTTGTATGAGAAACCCCACTCTGATATGAGGCC 541
Db 141592 CTCATTACATGAGACCTTCTATTACATGAGATCAGCTCTTATTACATGAGAC 141533
Cy 542 CACTCTCGGATATGAGAGCCCACTGAGAGAAATGAGAGCCCGCTGAGATACAGAG 601
Db 141532 CTCTCTATTACATGAGATCAGCTCTTATTACATGAGAGACCTCTTATTACATGAGAT 141473
Cy 602 CCTCACCTGCTGATGAGAGAGCAGGCTTCAGGATCTACAGCA 645
Db 141472 CACTCTCTTATTACATGAGACCTTCTTATTACAGAGCA 141429

RESULT 15

LOCUS

AC108000 167587 bp DNA linear PRI 21-MAY-2002
Homo sapiens chromosome 15, clone CTD-2116G1, complete sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

HTG.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 167587)

Bliren, B., Linton, L., Nusbaum, C., and Lander, B.

Unpublished

2 (bases 1 to 167587)

Bliren, B., Linton, L., Nusbaum, C., Lander, B., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouhgalter, B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., DeArrelano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S.,

Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, J., Grand-Pierre, N.,

Hagoe, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., Labocque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,

Mackdonald, P., Major, J., Margis, N., Matthews, C., McCarthy, M.,

McBarn, P., McKernan, K., Melidim, J., Meneus, L., Mihova, T.,

Mlena, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,

Peterson, K., Phunthang, P., Plerre, N., Pollare, V., Raymond, C.,

Retta, R., Riebeck, M., Riley, R., Rise, C., Rogov, P., Roman, J.,

Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Straus, N., Subramanian, A., Talama, J., Teste, S., Theodore, J.,

Topham, K., Travers, M., Travis, N., Trifillo, J., Vassiliev, H.,

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,

Zainoun, J., Zember, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 167587)

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,

Bouhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., DeArrelano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, J., Grand-Pierre, N.,

Hagoe, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menais, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., Pollara, V., Oliver, J., Peterson, K., Phunhthang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, K., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 167587)

TITLE

JOURNAL

COMMENT

Submitted (21-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 21, 2002 this sequence version replaced gl:20336146.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seg.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: L24580

Center clone name: 2116_G1

FEATURES

source Location/Qualifiers

1. 167587
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="CTD-2116G1"
/clone_lib="CITD1 Human BAC"
repeat_region complement(11..144)
repeat_region /rpt_family="L2"
repeat_region complement(145..356)
repeat_region /rpt_family="AluSg/X"
repeat_region complement(357..1109)
repeat_region /rpt_family="L2"
repeat_region 3713..3828
repeat_region /rpt_family="L1MB"

repeat_region complement(3834..4006)
repeat_region /rpt_family="MIR"
repeat_region 4574..4603
repeat_region /rpt_family="(GAAA)n"
repeat_region complement(4976..5167)
repeat_region /rpt_family="MIR"
repeat_region 5639..5775
repeat_region /rpt_family="L1"
repeat_region 6745..6766
repeat_region /rpt_family="AT-rich"
repeat_region complement(8290..8370)
repeat_region /rpt_family="MIR3"
repeat_region 9077..9288
repeat_region /rpt_family="MIR"
repeat_region 9496..10838
repeat_region /rpt_family="L1MD1"
repeat_region 10839..11162
repeat_region /rpt_family="AluSx"
repeat_region 11163..12229
repeat_region /rpt_family="L1MD1"
repeat_region complement(13118..13414)
repeat_region /rpt_family="AluY"
repeat_region 13667..14061
repeat_region /rpt_family="L2"
repeat_region 14366..14395
repeat_region /rpt_family="AT-rich"
repeat_region 14748..15052
repeat_region /rpt_family="AluSx"
repeat_region complement(15603..15768)
repeat_region /rpt_family="MIR"
repeat_region complement(15878..15932)
repeat_region /rpt_family="L2"
repeat_region complement(15988..16069)
repeat_region /rpt_family="MER1"
repeat_region complement(16309..16442)
repeat_region /rpt_family="MER46A"
repeat_region complement(16557..16730)
repeat_region /rpt_family="MIR"
repeat_region complement(16746..16830)
repeat_region /rpt_family="MIR"
repeat_region 16845..16939
repeat_region /rpt_family="MER117"
repeat_region complement(17267..17436)
repeat_region /rpt_family="MIR3"
repeat_region complement(17914..18014)
repeat_region /rpt_family="MER103"
repeat_region complement(19124..19415)
repeat_region /rpt_family="AluJo"
repeat_region complement(19571..19636)
repeat_region /rpt_family="L2"
repeat_region 19637..19926
repeat_region /rpt_family="AluJo"
repeat_region complement(19927..19984)
repeat_region /rpt_family="L2"
repeat_region complement(20449..21788)
repeat_region /rpt_family="L2"
repeat_region 21985..22144
repeat_region /rpt_family="MIR"
repeat_region 22728..22838
repeat_region /rpt_family="Char11e8"
repeat_region 23014..23044
repeat_region /rpt_family="AT-rich"
repeat_region 23140..23356
repeat_region /rpt_family="FLAM-C"
repeat_region 23265..23350
repeat_region /rpt_family="(TA)n"
repeat_region 23357..23381
repeat_region /rpt_family="AT-rich"
repeat_region 24908..25306
repeat_region /rpt_family="AluSx"
repeat_region complement(25805..25979)
repeat_region /rpt_family="L1MB4A"
repeat_region complement(26043..26262)

```
/rpt_family="L1ME4A"  
repeat_region complement(26861..26993)  
/rpt_family="PLAM C"  
repeat_region complement(27023..27317)  
/rpt_family="AluSx"  
repeat_region 27952..28300
```

Query Match 16.1%; Score 113.8; DB 9; Length 167587;

Best Local Similarity 55.1%; Pred. No. 1.6e-18;

Matches 223; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

```
QY 297 TGTCTATGGGGCCCCACCTGACGATATGAGCCCCACCTCCGAGATACGAGCCCCACC 356  
Db 119174 TGTTCACACAGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAACAGCCTCTCC 119233  
QY 357 TGCAGGATATGAGAGCCCAACCCGAGAAATGAAGGCCCGCTGTGGATACAGAGCTC 416  
Db 119234 TCCGTTCACGTAGCCTCTCTCTGTTCAACAGCCTCTCTCTGTTCAACGTAGCCTC 119293  
QY 417 ACCGTGCGATATGAGAGCCCAACCTTGTGATACGAGCCCACTGACAGATATGAGC 476  
Db 119294 TCCCTCTGTTCAACAGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAACAGC 119353  
QY 477 CCCACCTCTAGGATATGAGAGCCCACTCTTGTGATATGAACCCACCTCTCGGATATG 536  
Db 119354 CTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAAGT 119413  
QY 537 AGCCCCACCTCTCGGATATGAGAGCCCACTGACAGAAATGAAGGCCCGCTGCGGAGTA 596  
Db 119414 AGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCA 119473  
QY 597 CAGAGCTCACTGCTGATCAAGGACAGGCTCAGGAATCTACAGAGCCCAAGGCTCC 656  
Db 119474 CACAGCTCTCTCTCTGTTCAAGTAGCCTCTCTCTGTTCAACAGCCTCTCTCTGTT 119533  
QY 657 TGAAGAGAGGCTCTCTCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 701  
Db 119534 TCATGTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 119578
```

Search completed: December 16, 2003, 15:10:22
Job time : 2626.03 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 05:54:30 ; Search time 188.14 Seconds

(Without alignments)
10115.338 Million cell updates/sec

Title: US-09-864-291-11_COPY_1_705

Perfect score: 705
1 atgcattgatcgtatgcac.....cctcttcagatccatctc 705

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	705	100.0	1001	24	AA520602	DNA encoding human
2	446.2	63.3	467	23	AA101229	Human reproductive
3	446.2	63.3	467	23	AB196688	Human testicular a
4	414.4	58.8	220895	24	ABK84798	Human cDNA differ
5	411.4	58.4	436	22	ABA67850	Human foetal liver
6	411.4	58.4	436	22	AAK42003	Human bone marrow
7	411.4	58.4	436	22	AA148070	Probe 16756 used
8	411.4	58.4	436	24	AB516034	Human genome-deriv

9	385	54.6	1413	24	AA520601	DNA encoding bovin
10	375.4	53.2	7099	22	AA104882	Human reproductive
11	375.4	53.2	7099	23	AB197776	Human testicular a
12	173.6	24.6	894	23	AB70582	DNA encoding novel
13	130.4	18.5	471	22	ABA55246	Human foetal liver
14	130.4	18.5	471	22	AAK28957	Human bone marrow
15	130.4	18.5	471	22	AA134906	Probe #3592 used t
16	130.4	18.5	471	22	AB503491	Human genome-deriv
17	110.2	15.6	321	24	AA101355	Human reproductive
18	110.2	15.6	321	23	AB196808	Human testicular a
19	104.2	14.8	2772	23	AA572787	DNA encoding novel
20	104.2	14.8	2772	23	ABA90525	DNA encoding novel
21	102.6	14.6	2243	22	ABA08657	Human extensin hom
22	100	14.2	1824	23	AA581488	DNA encoding novel
23	100	14.2	2850	23	AA579695	DNA encoding novel
24	93.2	13.2	291	22	AA104883	Human reproductive
25	93.2	13.2	291	23	AB197777	Human testicular a
26	89.4	12.7	1964	22	ABA45936	Human breast cell
27	89.4	12.7	1964	22	ABA56461	Human foetal liver
28	89.4	12.7	1964	22	ABA26096	Probe #4562 for ge
29	89.4	12.7	1964	22	AAK04622	Human brain expres
30	89.4	12.7	1964	22	AAK30133	Human bone marrow
31	89.4	12.7	1964	22	AA114731	Probe #4664 for ge
32	89.4	12.7	1964	22	AA136097	Probe #4783 used t
33	89.4	12.7	1964	22	AA104535	Probe #4526 used t
34	89.4	12.7	1964	23	AB529782	Human liver single
35	89.4	12.7	1964	24	AB504714	Human genome-deriv
36	87	12.3	543	22	ABA51054	Human breast cell
37	87	12.3	543	22	ABA69042	Human foetal liver
38	87	12.3	543	22	ABA55988	Probe #1454 for g
39	87	12.3	543	22	AAK17357	Human brain expres
40	87	12.3	543	22	AAK43154	Human bone marrow
41	87	12.3	543	22	AA123325	Probe #13858 for g
42	87	12.3	543	22	AA149232	Probe #17918 used
43	87	12.3	543	22	AA109525	Probe #9516 used t
44	87	12.3	543	23	AB542787	Human liver single
45	87	12.3	543	24	AB517230	Human genome-deriv

ALIGNMENTS

AA520602	standard; cDNA; 1001 BP.
AA520602	
09-APR-2002	(first entry)
DNA encoding human testicular WW domain binding protein (htbWP).	
Testicular WW domain binding protein; WBPH; perinuclear choca 32;	
PT32; contriceptive; fertility; oocyte activation; vaccine;	
globocontermy; spermiogenesis; spermatozoa; tyrosine kinase; c-Yes;	
immunontrapeptive; human; gene; ss.	
Homo sapiens.	
Key	Location/Qualifiers
primer_bind	1..18
/*tag= a	/note= "Binds forward primer AA520605"
/*tag= b	/product= "htbWP"
/*tag= c	/note= "Human testicular WW domain binding protein"
/*tag= c	/complement (984..1001)
/note= "Binds reverse primer AA520606"	
MO200190185-A2.	

PD 29-NOV-2001.
 XX
 PF 25-MAY-2001; 2001WO-CA00738.
 XX
 PR 25-MAY-2000; 2000CA-2307128.
 PR 25-MAY-2000; 2000US-206979P.
 XX
 PA (TOOH) UNIV QUEBENS KINGSTON.
 PA (OYOR-) UNIV OREGON HEALTH SCT.
 XX
 PI Olo R, Surovsky P;
 XX
 DR WPI: 2002-097644/13.
 DR P-PSDB; AAU74610.
 XX
 PT Isolated perinuclear theca 32 polypeptide that interacts with activated
 PT tyrosine kinase c-Yes, for enhancing fertility, creating/diagnosing
 PT diminished fertility and abnormal spermiogenesis and for providing
 PT contraception -
 XX
 PS Claim 62; Fig 4B; 103pp; English.
 XX
 CC The invention describes an isolated perinuclear theca 32 (PT32)
 CC polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is
 CC useful for: enhancing fertility in a mammal; treating globozoosperm, by
 CC expressing (I) in spermatozoa; inhibiting fertilization, by introducing
 CC (I) or its antigenic fragment into a mammal to elicit an immune
 CC response; enhancing the ability of round spermatozoa to activate oocytes;
 CC treating or diagnosing diminished fertility and abnormal spermiogenesis;
 CC in providing contraceptive agents; identifying contraceptive and
 CC fertility-enhancing agents. The polynucleotide is useful for producing
 CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
 CC for chromosome identification. An antibody against (I) is useful in
 CC immunological assays, in immunoneutralization methods, to identify cells
 CC expressing (I), and to purify (I) by affinity chromatography. A
 CC transgenic animal is useful as an animal model for studying human
 CC fertility and reproductive biology, and for screening compounds to
 CC identify modulators of oocyte activation. The use of (I) prevents the
 CC entry of components which are detrimental to embryonic development into
 CC the oocyte during oocyte activation with crude sperm extract and avoids
 CC the propagation of viruses such as HIV (human immunodeficiency virus) and
 CC SIV (simian immunodeficiency virus) carried in the sperm. This sequence
 CC encodes the human testicular WW domain binding protein (hWWBP), described
 CC in the method of the invention.
 CC
 XX
 SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
 Query Match 100.0%; Score 705; DB 24; Length 1001;
 Best Local Similarity 100.0%; Pred. No. 2.2e-200; Indels 0; Gaps 0;
 Matches 705; Conservative 0; Mismatches 0;

Db 301 TATGGGGCCCCACCTGACGATATGAGCCCACTCCGAGATAGAGCCCACTGCA 360
 Qy 361 GGATATGAGGCCCAACCCCTGTAGAAATGAGGCCCGCTGTGGATATAGAGCCCTACT 420
 Db 361 GGATATGAGGCCCAACCCCTGTAGAAATGAGGCCCGCTGTGGATATAGAGCCCTACT 420
 Qy 421 GTGCGATATGAGGCCCACTCTTGTATAGAGAGCCCACTGTGAGATATGAGCCCA 480
 Db 421 GTGCGATATGAGGCCCACTCTTGTATAGAGAGCCCACTGTGAGATATGAGCCCA 480
 Qy 481 CCTTAGATATGAGGCCCACTCTTGTATAGAGAGCCCACTCTTGTATAGAGCC 540
 Db 481 CCTTAGATATGAGGCCCACTCTTGTATAGAGAGCCCACTCTTGTATAGAGCC 540
 Qy 541 CCACTCTGTGATATGAGAGCCCACTGTGAGAGAAATGAGGCCCGCTGTGGATATGAG 600
 Db 541 CCACTCTGTGATATGAGAGCCCACTGTGAGAGAAATGAGGCCCGCTGTGGATATGAG 600
 Qy 601 GCCTGACCTGTGATATGAGAGCCCACTGTGAGAGAAATGAGAGCCCACTGTGAGAA 660
 Db 601 GCCTGACCTGTGATATGAGAGCCCACTGTGAGAGAAATGAGAGCCCACTGTGAGAA 660
 Qy 661 AACGAGGCTTCTCTTCCCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
 Db 661 AACGAGGCTTCTCTTCCCTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
 RESULT 2
 AAL01229
 ID AAL01229 standard; cDNA; 467 BP.
 XX
 AC AAL01229;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DE Human reproductive system related antigen cDNA SEQ ID NO: 1230.
 XX
 KW Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155320-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01339.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.

PR	14-AUG-2000	2000US-0225470
PR	14-AUG-2000	2000US-0225477
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226579
PR	22-AUG-2000	2000US-0226581
PR	22-AUG-2000	2000US-0226668
PR	23-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227189
PR	30-AUG-2000	2000US-0228984
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231443
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0232081
PR	14-SEP-2000	2000US-0233397
PR	14-SEP-2000	2000US-0233398
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0233399
PR	14-SEP-2000	2000US-0234401
PR	14-SEP-2000	2000US-0234401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234274
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237039
PR	13-OCT-2000	2000US-0239935
PR	13-OCT-2000	2000US-0239935
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241788
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527

PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251865.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HDMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-465570/50.
 DR P-PSDB; AAM95259.
 XX
 XX Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 PR
 XX
 PS Claim 1; SEQ ID NO 1230; 1297bp + Sequence Listing; English.
 CC
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a coding sequence of the
 CC invention.
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;

Query Match	Similarity	Score	446.2; DB: 22;	Length	467;
Best Local	Similarity	98.5%;	Pred. No. 3.8e-123;		
Matches	459;	Conservative	1;	Mismatches	5;
				Indels	1;
				Gaps	1
Qy	129	CAGAAATGAGATGCCATTGAAATTTGGCCAGTTGATGATGAAAGCTGCTGCTGTGC	188		
Db	2	CAGAAATGAGATGCCATTGAAATTTGGCCAGTTGATGATGAAAGCTGCTGCTGTGC	61		
Qy	189	CCGAGATTTCCACTTGAACCTTAATGATGATGTTGAGCTCTATGGAAATTTATGTAAT	248		
Db	62	CCGAGATTTCCACTTGAACCTTAATGATGATGTTGAGCTCTATGGAAATTTATGTAAT	121		
Qy	249	TACTGGGAAAGGATATGTGCACTCCACAGATGCTGTATCACTATATGTTATGGGCT	308		

DB 122 TACTGGGAGGAAATATGTGCACTCCAGACATGCTTGTTCAGTTATGTCTATGAGC 181
QY 309 CCACCTGAGAGATATGAGCCCACTCCGGATACGAGCCCACTGCGAGATATG 368
DB 182 CCACCTGAGATATGAGCCCACTCCGGATACGAGCCCACTGCGAGATATG 241
QY 369 AGCCCAACCCGTAATGAGAGCCCGCTGTGGATACAGAGCTTCACTGTGGATA 428
DB 242 AGCCCAACCCGTAATGAGAGCCCGCTGTGGATACAGAGCTTCACTGTGGATA 301
QY 429 TGGAGCCCACTGTAATGAGAGCCCACTGCGATATGAGCCCACTGCTTAG 488
DB 302 TGGAGCCCACTGTAATGAGAGCCCACTGCGATATGAGCCCACTGCTTAG 361
QY 489 ATATGAGAGCCCACTGTAATGAGAGCCCACTGCGATATGAGCCCACTGCTTAG 547
DB 362 ATATGAGAGCCCACTGTAATGAGAGCCCACTGCGATATGAGAGCCCACTGCTTAG 421
QY 548 TGGATATGAGAGCCCACTGCGAGAAATGAGAGCCCGCTGTGGG 593
DB 422 TGGATATGAGAGCCCACTGCGAGAAATGAGAGCCCGCTGTGGG 467

RESULT 3
ABL96688
ID ABL96688 standard; cDNA; 467 BP.
XX
AC ABL96688;
XX
XX 21-JUN-2002 (first entry)
DT
XX
DE Human testicular antigen encoding cDNA SEQ ID NO: 356.
XX
XX
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
XX
PN MO200155317-A2.
PD
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214866.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216860.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0232404.
PR 14-SEP-2000; 2000US-0232405.
PR 14-SEP-2000; 2000US-0232422.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 12-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483232/52.
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer -
XX
XX Claim 1; SEQ ID NO 356; 766bp; English.
XX
CC The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a cDNA of the
CC invention.
XX
XX Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
SQ
Query Match 63.3%; Score 446.2; DB 23; Length 467;
Best local similarity 98.5%; Pred. No. 3.8e-123;
Matches 455; Conservative 1; Mismatches 5; Indels 1; Gaps 1;
QY 129 CAGAAATGAGATGCCATTGATTTGCCAGTTGATGATGAAAGCTGCTGCTGCTGC 188
DB 2 CAGAAATGAGATGCCATTGATTTGCCAGTTGATGATGAAAGCTGCTGCTGCTGC 61
QY 189 CCGAGATTTCATTGAAAGCTTAAATGATGATGATGATGATGATGATGATGAT 248
DB 62 CCGAGATTTCATTGAAAGCTTAAATGATGATGATGATGATGATGATGATGAT 121
QY 249 TATCGGGAGAGGATATGTCACCTCCACAGATGCTGCTGCTGCTGCTGCTGCTG 308
DB 122 TATCGGGAGAGGATATGTCACCTCCACAGATGCTGCTGCTGCTGCTGCTGCTG 181

QY 309 CCNACCTGAGATATGAGAGCCCACTCCCGATAGAGAGCCCACTGAGATATG 368
DB 182 CCNACCTGAGATATGAGAGCCCACTCCCGATAGAGAGCCCACTGAGATATG 241
QY 369 AGCCCAACCCGTGAAATGAGAGCCCGCTGTGATGATGAGAGCTGCTGCTG 428
DB 242 AGCCCAACCCGTGAAATGAGAGCCCGCTGTGATGATGAGAGCTGCTGCTG 301
QY 429 TGAAGCCCACTGATGATGAGAGCCCACTGAGATATGAGAGCCCACTGAG 488
DB 302 TGAAGCCCACTGATGATGAGAGCCCACTGAGATATGAGAGCCCACTGAG 361
QY 489 ATATGAGAGCCCACTGATGATGAGAGCCCACTGAGATATGAGAGCCCACTG 547
DB 362 ATATGAGAGCCCACTGATGATGAGAGCCCACTGAGATATGAGAGCCCACTG 421
QY 548 TCGATATGAGAGCCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 593
DB 422 TCGATATGAGAGCCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 467
RESULT 4
ABK84798
ID ABK84798 standard; cDNA; 220895 BP.
XX
AC ABK84798;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1369.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; AIDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN W0200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US30821.
XX
PR 03-OCT-2000; 2000US-237189P.
XX
PA (GENE-) GENE LOGIC INC.
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
XX WPI; 2002-435328/46.
XX
DR
XX
XX Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
XX Claim 1; SEQ ID NO 1369; 114pp; English.
XX
PS
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) Gs by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a


```

RESULT 6
AAK42003
ID AAK42003 standard; DNA: 436 BP.
AC
XX AAK42003;
AC
XX 06-NOV-2001 (first entry)
DT
XX Human bone marrow expressed single exon probe SEQ ID NO: 16560.
DE
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
XX MO200157276-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001MO-US00668.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX
XX Example 4; SEQ ID NO: 16560; 658bp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
XX
XX Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
XX
Query Match 58.4%; Score 411.4; DB 22; Length 436;
Best Local Similarity 99.8%; Pred. No. 9.3e-113;
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 293 TTATTGCTATGAGGCCCCCACTGCAAGATATGAGGCCCACTCCCGATACGAGCCC 352
DB 1 TTATTGCTATGAGGCCCCCACTGCAAGATATGAGGCCCACTCCCGATACGAGCCC 60
QY 353 CACCTGAGATATGAGGCCCACTGCAAGATATGAGGCCCGCTGAGATACAGAG 412
DB 61 CACCTGAGATATGAGGCCCACTGCAAGATATGAGGCCCGCTGAGATACAGAG 120
QY 413 CCTCACTGTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGATATG 472
DB 121 CCTCACTGTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGATATG 180
QY 473 GAGGCCCACTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGATATG 532
DB 181 GAGGCCCACTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGATATG 240
QY 533 ATGAGGCCCACTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGATATG 592
DB 241 ATGAGGCCCACTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGATATG 300

```

```

QY 593 GATACAGAGCCTCAGCTGTGATCAGAGCCAGGCTCAGATATACAGAGCCAGG 652
DB 301 GATACAGAGCCTCAGCTGTGATCAGAGCCAGGCTCAGATATACAGAGCCAGG 360
QY 653 CTCCTGAAAACGAGGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 705
DB 361 CTCCTGAAAACGAGGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 413
RESULT 7
AAI48070
ID AAI48070 standard; DNA: 436 BP.
AC
XX AAI48070;
AC
XX 17-OCT-2001 (first entry)
DT
XX Probe #16756 used to measure gene expression in human placenta sample.
DE
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200157272-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001MO-US00663.
PF
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
XX Claim 25; SEQ ID NO 16756; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP).
XX The present sequence is one such probe. The probes are useful for
XX predicting a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
XX
XX Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
XX
Query Match 58.4%; Score 411.4; DB 22; Length 436;
Best Local Similarity 99.8%; Pred. No. 9.3e-113;
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 293 TTATTGCTATGAGGCCCCCACTGCAAGATATGAGGCCCACTCCCGATACGAGCCC 352
DB 1 TTATTGCTATGAGGCCCCCACTGCAAGATATGAGGCCCACTCCCGATACGAGCCC 60
QY 353 CACCTGAGATATGAGGCCCACTGCAAGATATGAGGCCCGCTGAGATACAGAG 412
DB 61 CACCTGAGATATGAGGCCCACTGCAAGATATGAGGCCCGCTGAGATACAGAG 120
QY 413 CCTCACTGTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGATATG 472
DB 121 CCTCACTGTGCGATATGAGGCCCACTGCGATATGAGGCCCACTGCGATATG 180

```

QY 473 GAGCCCCACCTCTAGATATGAGAGCCCACTCTTGGATATGAAACCCACCTCTGGAT 532
 DB 181 GAGCCCCACCTCTAGATATGAGAGCCCACTCTTGGATATGAAACCCACCTCTGGAT 240
 QY 533 ATGAGAGCCCACTCTCTGGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGCTGGG 592
 DB 241 ATGAGAGCCCACTCTCTGGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGCTGGG 300
 QY 593 GATACAGAGCCCTCACTCTCTGGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGAG 652
 DB 301 GATACAGAGCCCTCACTCTCTGGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGAG 360
 QY 653 CTCTGAGAGAGAGGCTTCTCTCTCTGAGCTCTCTCTCTGAGTCAATTC 705
 DB 361 CTCTGAGAGAGAGGCTTCTCTCTCTCTCTCTCTCTCTCTGAGTCAATTC 413
 RESULT 8
 ABS16034
 ID ABS16034 standard; DNA; 436 BP.
 AC ABS16034;
 DT 19-AUG-2002 (first entry)
 XX Human genome-derived single exon probe ORF from lung SRQ ID No 16025.
 DE Human genome-derived single exon probe ORF from lung SRQ ID No 16025.
 KW Human; de; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlik syndrome; sarcoidosis; pulmonary hemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.
 OS Homo sapiens.
 PN WO200186003-A2.
 PD 15-NOV-2001.
 PF 30-JAN-2001; 2001MO-US00665.
 XX 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-060840P.
 PR 03-AUG-2000; 2000US-063236P.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2002-114183/15.
 DR Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 PS Claim 4; SRQ ID No 16025; 634p; English.
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a

CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridization of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridization to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlik syndrome, sarcoidosis, pulmonary
 CC hemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karsagen syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPRO at
 CC ftp.wipro.int/pub/published_pct_sequences.
 XX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 58.4%; Score 411.4; DB 24; Length 436;
 Best Local Similarity 99.8%; Pred. No. 9.3e-113;
 Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 293 TTATTGTCTATGAGGCCCCCACTGAGAGATATGAGAGCCCACTCCCGATACGAGCC 352
 DB 1 TTATTGTCTATGAGAGCCCACTGAGAGATATGAGAGCCCACTCCCGATACGAGCC 60
 QY 353 CACCTGAGAGATATGAGAGCCCACTGAGAGAAATGAGAGCCCGCTGAGATACAGAG 412
 DB 61 CACCTGAGAGATATGAGAGCCCACTGAGAGAAATGAGAGCCCGCTGAGATACAGAG 120
 QY 413 CCTCACTGTGAGATATGAGAGCCCACTTGTGATACGAGAGCCCACTGAGATATG 472
 DB 121 CCTCACTGTGAGATATGAGAGCCCACTTGTGATACGAGAGCCCACTGAGATATG 180
 QY 473 GAGCCCCACCTCTAGATATGAGAGCCCACTCTTGGATATGAAACCCACCTCTGGAT 532
 DB 181 GAGCCCCACCTCTAGATATGAGAGCCCACTCTTGGATATGAAACCCACCTCTGGAT 240
 QY 533 ATGAGAGCCCACTCTCTGGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGCTGGG 592
 DB 241 ATGAGAGCCCACTCTCTGGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGCTGGG 300
 QY 593 GATACAGAGCCCTCACTCTCTGGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGAG 652
 DB 301 GATACAGAGCCCTCACTCTCTGGATATGAGAGCCCACTCTGAGAGAAATGAGAGCCCGAG 360
 QY 653 CTCTGAGAGAGAGGCTTCTCTCTCTGAGCTCTCTCTCTGAGTCAATTC 705
 DB 361 CTCTGAGAGAGAGGCTTCTCTCTCTCTCTCTCTCTCTGAGTCAATTC 413
 RESULT 9
 AAS20601
 ID AAS20601 standard; cDNA; 1413 BP.
 XX AAS20601;
 AC AAS20601;

XX Homo sapiens.
OS
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216447.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AGU-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-465570/50.
DR
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 7570; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
Query Match 53.2%; Score 375.4; DB 22; Length 7099;
Best Local Similarity 99.7%; Pred. No. 2.1e-101;
Matches 376; Conservative 1; Indels 0; Gaps 0;
QY 290 CAGTTATGTCATGAGGCCCCCAGCTGAGATATGAGCCCCCAGCTCCGAGATACGAG 349
DB 6723 CAGTTATGTCATGAGGCCCCCAGCTGAGATATGAGCCCCCAGCTCCGAGATACGAG 6782
QY 350 CCCCAGCTGAGATATGAGCCCCCAGCTGAGATATGAGCCCCCAGCTCCGAGATACGAG 409
DB 6783 CCCCAGCTGAGATATGAGCCCCCAGCTGAGATATGAGCCCCCAGCTCCGAGATACGAG 6842
QY 410 GAGGCTGAGCTGAGATATGAGCCCCCAGCTGAGATATGAGCCCCCAGCTCCGAGATACGAG 469
DB 6843 GAGGCTGAGCTGAGATATGAGCCCCCAGCTGAGATATGAGCCCCCAGCTCCGAGATACGAG 6902
QY 470 ATGAGGCCCACTCTAGATATGAGCCCCCAGCTCTTGAATGAGAACCCCACTCTCG 529
DB 6903 ATGAGGCCCACTCTAGATATGAGCCCCCAGCTCTTGAATGAGAACCCCACTCTCG 6962
QY 530 GATATGAGGCCCACTCTGAGATATGAGCCCCCAGCTGAGATATGAGGCCCGCTG 589
DB 6963 GATATGAGGCCCACTCTGAGATATGAGCCCCCAGCTGAGATATGAGGCCCGCTG 7022
QY 590 CGGATATCAGAGCTCTGAGATATGAGGCCCGCTGAGATATCAGAGAGCC 649
DB 7023 CGGATATCAGAGCTCTGAGATATGAGGCCCGCTGAGATATCAGAGAGCC 7082
QY 650 AGGCTCTGAGAAAGAG 666
DB 7083 AGGCTCTGAGAAAGAG 7099
RESULT 11
ABL97776
ID ABL97776 standard; DNA; 7099 BP.
XX
XX ABL97776;
AC
XX
DT 21-JUN-2002 (first entry)
XX
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2428.
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KM reproductive system disorder; urinary system disorder; gene therapy;
KM cardiovascular disorder; respiratory disorder; neurological disorder;
KM gastrointestinal disease; infection; cytostatic; gene; ds.
XX
XX Homo sapiens.
OS
XX
XX MO20015317-A2.
PN
XX
PD 02-AUG-2001.
XX

PR 17-JAN-2001; 2001WO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232081.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barabsh SC, Ruben SM,
 XX
 DR WPI; 2001-483232/52.
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides,

PT useful for preventing, diagnosing and/or treating testicular cancer -
 XX
 XX Disclosure; SEQ ID NO 2428; 766pp; English.
 PS
 XX The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer.
 CC especially testicular cancer. The present sequence is a DNA encoding a
 CC protein fragment of the invention.
 CC
 XX
 SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
 Query Match 53.2%; Score 375.4; DB 23; Length 7099;
 Best Local Similarity 99.7%; Pred. No. 2.1e-101;
 Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 230 CAGTATATGTCATATGGGCCCCCACCCTGACAGATATGAGCCCCCACTCCGGATACGAG 349
 DB 6723 CAGTATATGTCATATGAGCCCCCACCCTGACAGATATGAGCCCCCACTCCGGATACGAG 6782
 QY 350 CCCCACCTGACAGATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGACAGAT 409
 DB 6783 CCCCACCTGACAGATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGACAGAT 6842
 QY 410 GAGCCTCACTGTGAGATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGACAGAT 469
 DB 6843 GAGCCTCACTGTGAGATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGACAGAT 6902
 QY 470 ATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGAC 529
 DB 6903 ATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGAC 6962
 QY 530 GATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGAC 589
 DB 6963 GATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGACAGATATGAGCCCCCACCCTGAC 7022
 QY 590 CGGATACAGAGCCTCACTGCTGATACAGAGCCAGGCTCAGGAATCTACAGAGCCC 649
 DB 7023 CGGATACAGAGCCTCACTGCTGATACAGAGCCAGGCTCAGGAATCTACAGAGCCC 7082
 QY 650 AGGCTCTGAAAAACGAG 666
 DB 7083 AGGCTCTGAAAAACGAG 7099
 RESULT 12
 AAS70582
 ID AAS70582 standard; cDNA; 894 BP.
 XX
 XX AAS70582;
 AC
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #6386.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 XX

PI Dmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR P-PSDB; ABG06395.
 XX
 PT New isolated polypeptide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 XX
 PS Claim 1, SEQ ID NO 6386, 103pp, English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 894 BP; 204 A; 221 C; 213 G; 256 T; 0 other;
 Query Match 24.6%; Score 173.6; DB 23; Length 894;
 Best Local Similarity 86.8%; Pred. No. 1.7e-41;
 Matches 191; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
 QY 448 TACGAGCCCACTGAGATATGAGCCCACTCTAGATATGAGCCCACTCTT 507
 DB 165 TATGAAGCCCGCTGTGGATACAGAGCCCTCCTGTGATATGAGCCCACTCTT 224
 QY 508 GGATATGAGAACCCCACTCTGAGATATGAGCCCACTCTGATATGAGCCCACTT 567
 DB 225 GGATATGAGAACCCCACTCTGAGATATGAGCCCACTCTGATATGAGCCCACTT 284
 QY 568 GCAGAAATGAAGCCCGCTGTGGATATGAGAGCTTCACTGCTGATCAGAGCCAG 627
 DB 285 GCAGAAATGAAGCCCGCTGTGGATATGAGAGCTTCACTGCTGATCAGAGCCAG 344
 QY 628 CCTCAGGAATCTACAGAGCCCACTCTGAAAAAGAG 667
 DB 345 CCTCAGGAATCTACAGAGAAAGAGTGTGCTCCAGAAAGCTTGG 384
 RESULT 13
 ID ABA55246
 ID ABA55246 standard; DNA; 471 BP.
 XX ABA55246;
 AC
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #3551.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; 88.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.

PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI: 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver -
 XX
 PS Claim 1, SEQ ID NO 3551; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;
 Query Match 18.5%; Score 130.4; DB 22; Length 471;
 Best Local Similarity 99.2%; Pred. No. 1e-28;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 290 CAGTATTTGCTATGAGGAGCCCACTGAGATATGAGCCCACTCCGATACGAG 349
 DB 340 CAGTATTTGCTATGAGGAGCCCACTGAGATATGAGCCCACTCCGATACGAG 399
 QY 350 CCCCACTGAGATATGAGGAGCCCACTGAGAAATGAAGCCCGCTGTGGATACA 409
 DB 400 CCCCACTGAGATATGAGGAGCCCACTGAGAAATGAAGCCCGCTGTGGATACA 459
 QY 410 GAGCTTCACCTG 421
 DB 460 GAGCTTCACCTG 471
 RESULT 14
 ID AAK28957
 ID AAK28957 standard; DNA; 471 BP.
 XX AAK28957;
 AC
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 3514.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukemia; lymphoma; myeloma; 88.
 XX
 OS Homo sapiens.
 XX
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human bone marrow -
 XX
 PS Example 4; SEQ ID NO: 3514; 658bp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may be used to improve diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 CC
 SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;

Query Match 18.5%; Score 130.4; DB 22; Length 471;
 Best Local Similarity 99.2%; Pred. No. 1e-28;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 290 CAGTTATTGTCTATGGGGCCCCCAGCTGCAAGATATGAGCCCCCACTCCCGGATACGGAG 349
 DB 340 CAGTTATTGTCTATGGAGCCCCCAGCTGCAAGATATGAGCCCCCACTCCCGGATACGGAG 399
 QY 350 CCCCACCTGCAAGATATGAGCCCCCACTGAGAAATGAAAGCCCGCTGTGGATAC 409
 DB 400 CCCCACCTGCAAGATATGAGCCCCCACTGAGAAATGAAAGCCCGCTGTGGATAC 459
 QY 410 GAGCCTCACCTG 421
 DB 460 GAGCCTCACCTG 471

RESULT 15
 ID AAI34906 standard; DNA; 471 BP.
 AC AAI34906;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #3592 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 XX genetic disorder; ss.
 OS Homo sapiens.
 XX
 PN W0200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-48897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID NO 3592; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENPs).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 CC
 SQ Sequence 471 BP; 120 A; 112 C; 82 G; 157 T; 0 other;

Query Match 18.5%; Score 130.4; DB 22; Length 471;
 Best Local Similarity 99.2%; Pred. No. 1e-28;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 290 CAGTTATTGTCTATGGGGCCCCCAGCTGCAAGATATGAGCCCCCACTCCCGGATACGGAG 349
 DB 340 CAGTTATTGTCTATGGAGCCCCCAGCTGCAAGATATGAGCCCCCACTCCCGGATACGGAG 399
 QY 350 CCCCACCTGCAAGATATGAGCCCCCACTGAGAAATGAAAGCCCGCTGTGGATAC 409
 DB 400 CCCCACCTGCAAGATATGAGCCCCCACTGAGAAATGAAAGCCCGCTGTGGATAC 459
 QY 410 GAGCCTCACCTG 421
 DB 460 GAGCCTCACCTG 471

Search completed: December 16, 2003, 11:00:51
 Job time : 189.14 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:55:15 ; Search time 49.6677 Seconds

(without alignments)
6265.145 Million cell updates/sec

Title: US-09-864-291-11_COPY_1_705

Perfect score: 1 atgcattgatcgtatgcac.....cctctctcagtcattct 705

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgm2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgm2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgm2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgm2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgm2_6/ptodata/2/ina/PTCUS.COMB.seq:*
6: /cgm2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.2	11.0	1885	4	US-09-484-970B-87 Sequence 87, Appl
2	76	10.8	981	2	US-08-841-349-15 Sequence 15, Appl
3	54.4	7.7	243	1	US-08-182-175A-56 Sequence 56, Appl
4	54.4	7.7	243	1	US-08-474-633A-74 Sequence 74, Appl
5	54.4	7.7	243	4	US-08-823-771-74 Sequence 74, Appl
6	54.4	7.7	243	5	PCT-US92-06412-56 Sequence 56, Appl
7	53.6	7.6	3833	5	US-08-917-320-18 Sequence 18, Appl
8	53.6	7.6	5931	3	PCT-US95-04611A-16 Sequence 16, Appl
9	53.6	7.6	5931	3	US-08-783-774-1 Sequence 1, Appl
10	53.6	7.6	5931	4	US-09-556-706B-1 Sequence 1, Appl
11	52	7.4	2188	1	US-07-865-662F-10 Sequence 10, Appl
12	52	7.4	2188	3	US-08-374-219B-10 Sequence 10, Appl
13	52	7.4	3489	2	US-08-728-323A-1 Sequence 1, Appl
14	52	7.4	3489	4	US-09-298-568-1 Sequence 1, Appl
15	52	7.4	3489	4	US-09-410-399-1 Sequence 1, Appl
16	52	7.4	32207	2	US-08-770-379-20 Sequence 20, Appl
17	52	7.4	32207	3	US-08-757-669A-20 Sequence 20, Appl
18	52	7.4	32207	4	US-09-230-371A-20 Sequence 20, Appl
19	50.2	7.1	604	4	US-09-370-838-175 Sequence 175, App
20	50	7.1	7785	2	US-08-276-967-1 Sequence 1, Appl
21	49.6	7.0	3222	4	US-08-714-741-39 Sequence 39, Appl
22	48.6	6.9	2633	4	US-09-149-476-51 Sequence 51, Appl
23	47.2	6.7	5661	4	US-08-938-105-2 Sequence 2, Appl
24	44.6	6.3	1562	4	US-09-620-312D-510 Sequence 510, App
25	44.6	6.3	2144	3	US-08-834-306-15 Sequence 15, Appl
26	44.6	6.3	2144	3	US-08-993-674A-15 Sequence 15, Appl
27	44.6	6.3	2144	4	US-09-256-976-15 Sequence 15, Appl

C	28	44.4	6.3	1235	2	US-08-557-309B-53	Sequence 53, Appl
	29	43.2	6.1	187	1	US-08-182-175A-90	Sequence 90, Appl
	30	43.2	6.1	187	1	US-08-474-633A-78	Sequence 78, Appl
	31	43.2	6.1	187	4	US-08-823-771-78	Sequence 78, Appl
	32	43.2	6.1	187	5	PCT-US92-06412-90	Sequence 90, Appl
	33	43	6.1	8310	3	US-08-870-126-11	Sequence 11, Appl
	34	43	6.1	8310	4	US-09-445-247-11	Sequence 11, Appl
	35	43	6.1	14985	4	US-08-652-972A-6	Sequence 6, Appl
	36	43	6.1	14985	5	PCT-US96-06231A-6	Sequence 6, Appl
	37	42.4	6.0	981	2	US-08-841-349-15	Sequence 15, Appl
	38	42.2	6.0	1251	4	US-09-252-991A-444	Sequence 44, App
	39	42.2	6.0	1362	4	US-09-252-991A-417	Sequence 417, App
	40	42	6.0	1995	1	US-08-425-069-3	Sequence 3, Appl
	41	42	6.0	1995	2	US-08-317-844B-3	Sequence 3, Appl
	42	42	6.0	6314	4	US-09-620-312D-98	Sequence 98, Appl
	43	41	5.8	1562	4	US-09-620-312D-510	Sequence 510, App
	44	40.6	5.8	1849	2	US-08-665-926-7	Sequence 7, Appl
	45	40.6	5.8	1849	3	US-08-740-223A-9	Sequence 9, Appl

ALIGNMENTS

```
RESULT 1
US-09-484-970B-87
Sequence 87, Application US/09484970B
Patent No. 6426186
GENERAL INFORMATION:
APPLICANT: Jones, Karen A.
APPLICANT: Volkmut, Wayne
APPLICANT: Walker, Michael G.
TITLE OR INVENTION: BONE REMODELING GENES
FILE REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 1885
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 232968.10CB1
US-09-484-970B-87

Query Match      11.0%; Score 77.2; DB 4; Length 1885;
Best Local Similarity 62.4%; Pred. No. 8.5e-14;
Matches 121; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY      1 ATGCCATTGATGATGATGACGAACCTCATCTGTTGAACACAGATATTGCTGCAAACTTC 60
        |||||
DB      274 ATCCCAATTTATCTCATGAAAGACTGTGAGATCAAGACGCCATATTGGTGCAACTAC 333

QY      61 ATTAAGGAACATTTACAGGAGCTCATATGTCGTGCGGAGAGACAACTACTTTAA 120
        |||||
DB      334 ATCAAGGAACACATGAAAGCGGAAGCGGAGGTGCTGGGAAGCTCTGCTTCTTCAAG 393

QY      121 TTAGTCTTCAGAAATGAGATGCAATTTGATGCCATGTAATGTAAGTAAGTGCCTCT 180
        |||||
DB      394 TTGACTTTCACGCGAGGCGGCGCATTTGACTTTCAGACAGGAGATCTCCAGTGGGACTT 453

QY      181 GCTGTGCGCGAG 194
        |||||
DB      454 CAAGCTCCAGAG 467

RESULT 2
US-08-841-349-15/c
Sequence 15, Application US/08841349B
Patent No. 5955594
GENERAL INFORMATION:
APPLICANT: MISHRA, LOPA
```



```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGEL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: E. coli
; CELL TYPE: DH5 alpha
; IMMEDIATE SOURCE:
; CLONE: 2-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..235
; OTHER INFORMATION: /function= "synthetic
; OTHER INFORMATION: storage protein
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "gsp"
; OTHER INFORMATION: /standard_name=
; OTHER INFORMATION: "7.7.7.7.7.8.9.8.9.5"
; US-08-474-633A-74

Query Match 7.7%; Score 54.4; DB 1; Length 243;
Best Local Similarity 53.2%; Pred. No. 2.9e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 302 ATGGGGCCCCCGATGATGAGCCCGACCTCCCGATACGAGCCCCCAGCTGAG 361
DB 2 ATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAG 61
QY 362 GATATGAGCCCAACCGTAGAATGAGAGCCCGCTGTGGATACAGAGCTCACCTG 421
DB 62 GCATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTG 121
QY 422 TGCATATGAGAGCCCGACCTTGTGATACGAGAGCCCGACCTGCGAGATATGAGAGCCCGAC 481
DB 122 AAGCGATGAGAGAGAGCTTAAGAGATGAGAGAGAGAGCTGAAGATGAGAGAGAGAA 181
QY 482 CTCTAGATATGAGAGCCCGACCTTGTGATATGAA 517
DB 182 CTCAGAAAGATGAGAGAGAGCTTAATGATGAGAA 217

RESULT 5
US-08-823-771-74
; Sequence 74, Application US/08823771
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; METHODS FOR INCREASING
; INCREASING THE LYSINE
; AND TREONINE CONTENT
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; AND COMPANY
```

```

; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/823,771
; FILING DATE: 24-Mar-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/474,633
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGEL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: E. coli
; CELL TYPE: DH5 alpha
; IMMEDIATE SOURCE:
; CLONE: 2-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..235
; OTHER INFORMATION: /function= "synthetic
; OTHER INFORMATION: storage protein
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "gsp"
; OTHER INFORMATION: /standard_name=
; OTHER INFORMATION: "7.7.7.7.7.8.9.8.9.5"
; SEQUENCE DESCRIPTION: SEQ ID NO: 74:
; US-08-823-771-74

Query Match 7.7%; Score 54.4; DB 4; Length 243;
Best Local Similarity 53.2%; Pred. No. 2.9e-07;
Matches 115; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 302 ATGGGGCCCCCGATGATGAGCCCGACCTCCCGATACGAGCCCCCAGCTGAG 361
DB 2 ATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAG 61
QY 362 GATATGAGCCCAACCGTAGAATGAGAGCCCGCTGTGGATACAGAGCTCACCTG 421
DB 62 GCATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTG 121
QY 422 TGCATATGAGAGCCCGACCTTGTGATACGAGAGCCCGACCTGCGAGATATGAGAGCCCGAC 481
DB 122 AAGCGATGAGAGAGAGCTTAAGAGATGAGAGAGAGAGCTGAAGATGAGAGAGAGAA 181
QY 482 CTCTAGATATGAGAGCCCGACCTTGTGATATGAA 517
DB 182 CTCAGAAAGATGAGAGAGAGCTTAATGATGAGAA 217

RESULT 6
PCT-US92-06412-56
; Sequence 56, Application PC/TUS9206412
```


RESULT 8
PCT-US95-04611A-18
Sequence 18, Application PC/US9504611A
GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: Non Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04611A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Luann Cseri
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/000US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
TELEX: 380816 CooleyPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
PCT-US95-04611A-18

Query Match 7.6%; Score 53.6; DB 5; Length 3833;
Best Local Similarity 50.4%; Pred. No. 2.3e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 307 GCCCACCCTGAGATATGAGGCCCACTCCCGATACGAGGCCCACTGAGATAT 366
DB 2578 GCCCACCCTGAGATATGAGGCCCACTCCCGATACGAGGCCCACTGAGATAT 2637
QY 367 GAGGCCCAACCCGTAGAAATGAGGCCCGCTGTGGATACAGAGCTCACTGTGCA 426
DB 2638 CTACCCCAACCCCAATGCAACGAGGCCCACTGTGGAAAAACAAGCTCACTGAG 2697
QY 427 TAGGAGCCCACTCTTGGATACGAGGCCCACTGAGATATGAGGCCCACTCTA 486
DB 2698 TGACTACCCCAACCCCAATGCAACGAGGCCCACTGTGGAAAAACAAGCTCACT 2757
QY 487 GATATGAGGCCCACTCTTGGATATGAGAACCCCACTCTGAGATATGAGGCCCACT 546
DB 2758 CAGTACTACCCCAACCCCAATGCAACGAGGCCCACTGTGGAAAAACAAGCTCACT 2817
QY 547 CTCGATATGAGGCCCACT 566
DB 2818 CAGCAGTACTACCCCACT 2837

RESULT 9
US-08-783-774-1
Sequence 1, Application US/08783774
Patent No. 6054130
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
APPLICANT: Jackman, Winthrop
TITLE OF INVENTION: NON-SPLICING VARIANTS OF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/783,774
FILING DATE: 15-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7682-037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1014..3734
OTHER INFORMATION:
US-08-783-774-1

Query Match 7.6%; Score 53.6; DB 3; Length 5931;
Best Local Similarity 50.4%; Pred. No. 2.9e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 307 GCCCACCCTGAGATATGAGGCCCACTCCCGATACGAGGCCCACTGAGATAT 366
DB 2578 GCCCACCCTGAGATATGAGGCCCACTCCCGATACGAGGCCCACTGAGATAT 2637
QY 367 GAGGCCCAACCCGTAGAAATGAGGCCCGCTGTGGATACAGAGCTCACTGTGCA 426
DB 2638 CTACCCCAACCCCAATGCAACGAGGCCCACTGTGGAAAAACAAGCTCACTGAG 2697
QY 427 TAGGAGCCCACTCTTGGATACGAGGCCCACTGAGATATGAGGCCCACTCTA 486
DB 2698 TGACTACCCCAACCCCAATGCAACGAGGCCCACTGTGGAAAAACAAGCTCACT 2757
QY 487 GATATGAGGCCCACTCTTGGATATGAGAACCCCACTCTGAGATATGAGGCCCACT 546
DB 2758 CAGTACTACCCCAACCCCAATGCAACGAGGCCCACTGTGGAAAAACAAGCTCACT 2817
QY 547 CTCGATATGAGGCCCACT 566
DB 2818 CAGCAGTACTACCCCACT 2837

RESULT 10

US-09-556-706B-1
Sequence 1, Application US/09556706B
Patent No. 6458364
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220
FILE REFERENCE: 7682-050-999
CURRENT APPLICATION NUMBER: US/09/556,706B
CURRENT FILING DATE: 2000-04-24
PRIOR APPLICATION NUMBER: 08/783,774
PRIOR FILING DATE: 1997-01-15
PRIOR APPLICATION NUMBER: 08/229,291
PRIOR FILING DATE: 1994-04-18
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 5931
TYPE: DNA
ORGANISM: Virus
FEATURE:
OTHER INFORMATION: gp350/220
US-09-556-706B-1

Query Match 7.6%; Score 53.6; DB 4; Length 5931;
Best Local Similarity 50.4%; Pred. No. 2.9e-06;
Matches 131; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
DB 307 GCCCAGCTGAGATATGAGAGCCCACTCCCGATGAGAGAGCCCACTTGCAGATAT 366
2578 GCGCCAGCCGAGAGTGAATGAGAGCCCAAAATGAGAGAGCCCACTTGCAGATAT 2637
QY 367 GAGAGCCCAAGGATGAGAGAGCCCGCTGAGATGAGAGAGCCCACTTGCAGATAT 426
DB 2638 CTAACCCCAAGGATGAGAGAGCCCGCTTGGAGAGAGAGCCCACTTGCAGATAT 2697
QY 427 TATGAGAGCCCAAGGATGAGAGAGCCCGCTTGGAGAGAGAGCCCACTTGCAGATAT 486
DB 2698 TGAATGAGAGAGCCCAAGGATGAGAGAGCCCGCTTGGAGAGAGAGCCCACTTGCAG 2757
QY 487 GATATGAGAGAGCCCAAGGATGAGAGAGCCCACTTGCAGATGAGAGAGCCCACTTGCAG 546
DB 2758 CAGTATGAGAGAGCCCAAGGATGAGAGAGCCCACTTGGAGAGAGAGAGCCCACTTGCAG 2817
QY 547 CTCGATATGAGAGAGCCCACTTGCAGATGAGAGAGCCCACTTGCAGATGAGAGAGCCCACT 566
DB 2818 CAGCAGTATGAGAGAGCCCACTTGCAGATGAGAGAGCCCACTTGCAGATGAGAGAGCCCACT 2837

RESULT 11
US-07-865-662F-10
Sequence 10, Application US/07865662F
Patent No. 5451670
GENERAL INFORMATION:
APPLICANT: Marcia M. Miller
TITLE OF INVENTION: Restriction Fragment Length
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: City of Hope
STREET: 1500 East Duarte Road
CITY: Duarte
STATE: California
COUNTRY: United States of America
ZIP: 91010-0269
COMPUTER READABLE FORM:
MEDIUM TYPE: 3M Double Density 5 1/4" diskette
COMPUTER: Wang PC
OPERATING SYSTEM: MS DOS Version 3.20
SOFTWARE: Microsoft
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/865,662F
FILING DATE: 07 April, 1992

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/688,326
FILING DATE: 22 April 1991
APPLICATION NUMBER: 07/588,922
FILING DATE: 27 September 1990
APPLICATION NUMBER: 07/210,405
FILING DATE: 23 June 1988
APPLICATION NUMBER: US 07/130,529
FILING DATE: 9 December 1987
APPLICATION NUMBER: US 07/068,176
FILING DATE: 30 June 1987
ATTORNEY/AGENT INFORMATION:
NAME: Irons, Edward S.
REGISTRATION NUMBER: 16,541
REFERENCE/DOCKET NUMBER: NO. 5451670e
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 783-6040
TELEFAX: (202) 783-6031
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE: Synthetically Prepared
IMMEDIATE SOURCE: Synthetically Prepared
US-07-865-662F-10

Query Match 7.4%; Score 52; DB 1; Length 2188;
Best Local Similarity 47.5%; Pred. No. 5.2e-06;
Matches 154; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
DB 302 ATGGGAGCCCACTGAGATATGAGAGCCCACTCCCGATGAGAGAGCCCACTTGCAGATAT 361
DB 1315 ATGGGAGCAAGCTTAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
QY 362 GATATGAGAGCCCAAGGATGAG 421
DB 1375 GAGATGAG 1434
QY 422 TGGATGAG 481
DB 1435 GAGAGATGAG 1494
QY 482 CTCTAGATATGAG 541
DB 1495 CATGAG 1554
QY 542 CAGCTCTGAGATATGAG 601
DB 1555 CAGCATGAG 1614
QY 602 CCTCAGCTGCTGATGAG 625
DB 1615 GAGCAGCATGAG 1638

RESULT 12
US-08-374-219B-10
Sequence 10, Application US/08374219B
Patent No. 6218106
GENERAL INFORMATION:
APPLICANT: Miller, Marcia M.
TITLE OF INVENTION: Restriction Fragment Length Polymorphism Test
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSER: Rothwell, F199, Ernst & Kurtz
STREET: 555 Thirteenth Street, N.W.
CITY: Washington

```

STATE: D. C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374,219B
FILING DATE: 18-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/865,662
FILING DATE: 07-APR-1992
APPLICATION NUMBER: 07/688,326
FILING DATE: 22-APR-1991
APPLICATION NUMBER: 07/588,922
FILING DATE: 27-SEP-1990
APPLICATION NUMBER: 07/413,301
FILING DATE: 28-SEP-1989
APPLICATION NUMBER: 07/210,405
FILING DATE: 23-JUN-1988
APPLICATION NUMBER: 07/130,529
FILING DATE: 09-DEC-1987
APPLICATION NUMBER: 07/068,176
FILING DATE: 30-JUN-1987
ATTORNEY/AGENT INFORMATION:
NAME: Walker, Barbara W
REGISTRATION NUMBER: 35,400
REFERENCE/DOCKET NUMBER: 2124-113A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202 783-6040
TELEFAX: 202 783-6031
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-374-2198-10

```

Query Match	7.4%	Score 52	DB 3	length 2188
Best Local Similarity	47.5%	Pred. No. 5.2e-06		
Matches 154	Conservative 0	Mismatches 170	Indels 0	Gaps 0

Query	302	ATGGGGCCCCCACTG	CAGATATGAGGCCCACTCCGGATACGAGGCCCACTTGAG	361
Db	1315	ATGGGGCAACAAGCTAAAGAAATCA	GAGGAACGAAATGGAGCTGAAGAGCACCATAG	1376
QY	362	GATATGAGGCCCAACCCGTTAGAAATGAAGCCCGCTGTGGATACAGAGCTTCACTG	421	
Db	1375	GAGATGGGGCAACAAGCTAAAGAAATCA	GAGGAACGAAATGGAGCTGAAGAGCACCAT	1434
QY	422	TGCGATATGAGGCCCACTCTTGATATCGAGGCCCACTTGACGATATGAGGCCCAAC	481	
Db	1435	GAGGAGATGGGGCAACAAGCTAAAGAAATCA	GAGGAACGAAATGGAGCTGAAGAGCACCAT	1492
QY	482	CTTAGATATGAGGCCCACTCTTGATATGAAACCCCACTCTCGGATATGAGGCC	541	
Db	1495	CATGAGGAGATGGGGCAACAAGCTAAAGAAATCA	GAGGAACGAAATGGAGCTGAAGAG	1552
QY	542	CACCTCTGGATATGAGGCCCACTTGACGAAATGAAGGCCCGCTCGGATACAGAG	601	
Db	1555	CACCTAGGAGATGGGGCAACAAGCTAAAGAAATCA	GAGGAACGAAATGGAGCTGAAG	1612
QY	602	CCTCACTGCTGGATCAGAGCCA	625	
Db	1615	GAGGACCATGAGAGATGGGGCA	1638	

RESULT	13
--------	----

```

US-08-728-323A-1
Sequence 1, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Ruess, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

```

Query Match	7.44	Score 52	DB 2	Length 3489
Best Local Similarity	48.1%	Prod. No. 6.7e-06		
Matches 148	Conservative	0	Mismatches 160	Indels 0
			Gaps 0	
QY	316	GCAGCATATGACCCCACTCCCGATACGAGACCCCACTGCAGGATATGAGCCCA	375	
DB	2202	GCACGACACAGAGATGACAGCAGCAGCATGACAGCAGCAGATGAACAGGA	2261	
QY	376	CCCGTAGAAATGAGGCCCGCCTGTGGGATACAGACCTCACCTGTGCATATGAGCC	435	
DB	2262	GCACGACGAGAGACGAGCAGCAGGAGGAGCAGAGCAGAGATTAGAGGACGAGCA	2321	
QY	436	CCACCTTTGGGATACGAGCCCCCACTGCAGATATGAGCCCCCACTTACGATATGGA	495	
DB	2322	GGAGTTAGAGGATCAGAGGAGGAGATTAGGAGGACAGACAGAGATTAGGAGCAGGA	2381	
QY	496	GCCCACTCTTGGATATGGAACCCCACTCGGATATGAGGCCCACTCTCGATAT	555	
DB	2382	GCAGAGTTAGAGGACGAGCAGGAGTTAGAGGACAGAGAGAGATTAGAGGAGCA	2441	
QY	556	GGAGCCCCCACTGCAGGAAATGAGGCCCGCCTGCAGGATACAGAGCTCACCTGTGGA	615	
DB	2442	GCAGCAGGAGTTAGAGGACGAGCAGGAGTTAGAGGACAGAGCAGAGATTAGAGGA	2501	
QY	616	TCAGGAGC	623	

Db 2502 GCAGGAGC 2509

RESULT 14

US-09-298-568-1 Application US/09298568

Sequence 1, Application US/09298568

Patent No. 6322792

GENERAL INFORMATION:

APPLICANT: Kieff, Elliott D.

APPLICANT: Bailestas, Mary E.

APPLICANT: Kaye, Kenneth M.

TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO

TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE

FILE REFERENCE: 16412-10001R

CURRENT APPLICATION NUMBER: US/09/298,568

EARLIER FILING DATE: 1999-04-21

EARLIER FILING DATE: 1998-11-19

NUMBER OF SEQ ID NOS: 3

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1

LENGTH: 3489

TYPE: DNA

ORGANISM: Kaposi's sarcoma-associated herpesvirus

US-09-298-568-1

Query Match 7.4%; Score 52; DB 4; Length 3489;

Best Local Similarity 48.1%; Pred. No. 6.7e-06;

Matches 148; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 316 GCAGGATATGAGACCCCACTCCCGATACGAGACCCCACTGAGATATGAGACCCCA 375
 Db 2202 GCAGGACGACGAGATGAGAGCAGCAGAGATGAGAGCAGAGATGAGAGCAGAG 2261
 QY 376 CCCGTAGGAATGAGGCCCGCTGTGGATACGAGACCTCAGCTGTGAGATGAGACC 435
 Db 2262 GCAGGACGAGAGAGCAGAGCAGAGAGCAGAGCAGAGAGCAGAGATGAGAGCAG 2321
 QY 436 CCACTCTTGGATACGAGACCCCACTGAGATGAGACCCCACTTAGAGATGAG 495
 Db 2322 GAGTTAGAGATCAGAGAGCAGAGATGAGAGCAGAGCAGAGATGAGAGCAGAG 2381
 QY 496 GCCCACTCTTGGATATGAGACCCCACTCTGAGATATGAGACCCCACTCTGAGAT 555
 Db 2382 GCAGGAGTTAGAGAGCAGAGCAGAGATGAGAGCAGAGCAGAGATGAGAGCAG 2441
 QY 556 GAGGCCCACTGAGAGGAATGAGGCCCGCTGCGGATACAGAGCCTCAGCTGTGAG 615
 Db 2442 GAGGACGAGATGAGAGCAGAGCAGAGATGAGAGCAGAGCAGAGATGAGAGCAG 2501
 QY 616 TCAGGAGC 623
 Db 2502 GCAGGAGC 2509

RESULT 15

US-09-410-399-1

Sequence 1, Application US/09410399

Patent No. 6482587

GENERAL INFORMATION:

APPLICANT: Robertson, Erle S.

APPLICANT: Colter, Murray A.

TITLE OF INVENTION: Methods to inhibit or enhance the binding of viral DNA

TITLE OF INVENTION: to Genomic Host DNA

FILE REFERENCE: UM-03778

CURRENT APPLICATION NUMBER: US/09/410,399

NUMBER OF SEQ ID NOS: 6

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1

LENGTH: 3489

TYPE: DNA

ORGANISM: Kaposi's sarcoma-associated herpesvirus

US-09-410-399-1

Query Match 7.4%; Score 52; DB 4; Length 3489;

Best Local Similarity 48.1%; Pred. No. 6.7e-06;

Matches 148; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 316 GCAGGATATGAGACCCCACTCCCGATACGAGACCCCACTGAGATATGAGACCCCA 375
 Db 2202 GCAGGACGACGAGATGAGAGCAGCAGAGATGAGAGCAGAGATGAGAGCAGAG 2261
 QY 376 CCCGTAGGAATGAGGCCCGCTGTGGATACGAGACCTCAGCTGTGAGATGAGACC 435
 Db 2262 GCAGGACGAGAGAGCAGAGCAGAGAGCAGAGCAGAGATGAGAGCAGAGATGAGAGCAG 2321
 QY 436 CCACTCTTGGATACGAGACCCCACTGAGATATGAGACCCCACTTAGAGATGAG 495
 Db 2322 GAGTTAGAGATCAGAGAGCAGAGATGAGAGCAGAGCAGAGATGAGAGCAGAG 2381
 QY 496 GCCCACTCTTGGATATGAGACCCCACTCTGAGATATGAGACCCCACTCTGAGAT 555
 Db 2382 GCAGGAGTTAGAGAGCAGAGCAGAGATGAGAGCAGAGCAGAGATGAGAGCAGAG 2441
 QY 556 GAGGCCCACTGAGAGGAATGAGGCCCGCTGCGGATACAGAGCCTCAGCTGTGAG 615
 Db 2442 GAGGACGAGATGAGAGCAGAGCAGAGATGAGAGCAGAGCAGAGATGAGAGCAGAG 2501
 QY 616 TCAGGAGC 623
 Db 2502 GCAGGAGC 2509

Search completed: December 16, 2003, 17:36:30

Job time : 50.6677 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 10:43:10 ; Search time 232.016 Seconds

(without alignments)
10099.015 Million cell updates/sec

Title: US-09-864-291-11_COPY_1_705

Perfect score: 705

Sequence: 1 atgcatttcctcgtatgac.....cctcttcctcgtatcctc 705

Scoring table:

IDENTITY NUC
Gap 10.0 ; Gapext 1.0

Searched: 2201672 seqs, 166179959 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_PUB_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	446.2	58.4	467	11	US-09-864-891-1230 Sequence 1230, Ap
2	411.4	53.2	436	9	US-09-864-761-13561 Sequence 31561, A
3	375.4	53.2	7099	11	US-09-764-891-7570 Sequence 7570, A
4	130.4	18.5	471	9	US-09-864-761-15029 Sequence 15029, A
5	110.2	15.6	321	11	US-09-764-891-1356 Sequence 1356, Ap
6	109	15.5	866	13	US-10-029-386-22839 Sequence 22839, A
7	100	14.2	1040	13	US-10-029-386-26033 Sequence 26033, A
8	93.2	13.2	291	11	US-09-764-891-7571 Sequence 7571, Ap
9	89.4	12.7	1984	9	US-09-864-761-4562 Sequence 4562, Ap
10	88.8	12.6	541	13	US-10-029-386-12333 Sequence 12333, A
11	87.8	12.5	511	13	US-10-029-386-24981 Sequence 24981, A
12	87	12.3	543	9	US-09-864-761-21308 Sequence 21308, A
13	77.2	11.0	432	11	US-09-918-995-33536 Sequence 33536, A
14	69.2	9.8	541	13	US-10-029-386-11265 Sequence 11265, A
15	67.8	9.6	511	13	US-10-029-386-24981 Sequence 24981, A

16	64.8	9.2	593	13	US-10-027-632-290832 Sequence 290832, A
17	64.8	9.2	593	14	US-10-027-632-290832 Sequence 290832, A
18	63.2	9.0	409	10	US-09-860-352-6984 Sequence 6984, Ap
19	62.2	8.8	689	13	US-10-027-632-148124 Sequence 148124, A
20	62.2	8.7	509	9	US-09-864-761-25366 Sequence 25366, A
21	61.2	8.7	596	9	US-09-864-761-8648 Sequence 8648, Ap
22	61.2	8.7	596	9	US-09-864-761-8648 Sequence 8648, Ap
23	57.4	8.1	2489	11	US-09-298-5238-72 Sequence 72, Ap1
24	57.2	8.1	436	9	US-09-864-761-13561 Sequence 31561, A
25	57.2	8.1	7099	11	US-09-764-891-7570 Sequence 7570, Ap1
26	54.4	7.7	243	15	US-10-023-0664-74 Sequence 74, Ap1
27	54.2	7.7	1218	9	US-09-864-761-20761 Sequence 20761, A
28	54.2	7.7	1954	9	US-09-864-761-4001 Sequence 4001, Ap
29	54.2	7.7	3125	15	US-10-175-523-151 Sequence 151, Ap
30	54.2	7.7	4768	10	US-09-842-777-9 Sequence 9, Ap11
31	54.2	7.7	5994	13	US-10-205-219-136 Sequence 136, Ap
32	53.2	7.5	4395	9	US-09-925-301-224 Sequence 224, Ap
33	53.2	7.5	8412	11	US-09-919-039-223 Sequence 223, Ap
34	52.8	7.5	2016	10	US-09-938-8428-2004 Sequence 2004, Ap
35	52.6	7.5	5327	13	US-10-240-453-64 Sequence 64, Ap1
36	52.6	7.5	5327	13	US-10-239-676-62 Sequence 62, Ap1
37	52	7.4	3489	15	US-10-294-804-1 Sequence 1, Ap11
38	50.2	7.1	604	10	US-09-738-973-175 Sequence 175, Ap
39	50.2	7.1	604	10	US-09-854-133-175 Sequence 175, Ap
40	50.2	7.1	604	15	US-10-144-6498-175 Sequence 175, Ap
41	50	7.1	895	9	US-09-864-761-23227 Sequence 23227, A
42	49.6	7.0	591	9	US-09-864-761-19727 Sequence 19727, A
43	49.6	7.0	1959	9	US-09-864-761-2946 Sequence 2946, Ap
44	49.6	7.0	3463	11	US-09-298-5238-59 Sequence 59, Ap1
45	48.8	6.9	5918	15	US-10-198-846-13875 Sequence 13875, A

ALIGNMENTS

RESULT 1

US-09-764-891-1230

Sequence 1230, Ap

Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

PRIOR REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

PRIOR FILING DATE: 2001-01-17

PRIOR APPLICATION DATA REMOVED - consult PAM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1230

LENGTH: 467

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

NAME/KEY: SITE

LOCATION: (421)

OTHER INFORMATION: n equals a,t,g, or c

US-09-764-891-1230

Query Match	63.3%	Score 446.2	DB 11	Length 467
Best Local Similarity	98.5%	Pred. No. 1.4e-136		
Matches 459	Conservative 1	Mismatches 5	Indels 1	Gaps 1
Qy	129	CAGAAATGAGATGCCATTGAATTCGCGAGTGTGATGATGAAAGCTGCTGCTGTCG 188		
Db	2	CAGAAATGAGATGCCATTGAATTCGCGAGTGTGATGATGAAAGCTGCTGCTGTCG 61		
Qy	189	CCGAGATTTTCATTGAACCTTAATGACTGTTGAGCTCTAAGGAAATTATGTAAT 248		
Db	62	CCGAGATTTTCATTGAACCTTAATGACTGTTGAGCTCTAAGGAAATTATGTAAT 121		
Qy	249	TACTGGGAGGAGGAAATATGCACTCAGAGATGCTTGTACGTTATGTTATGAGGGC 308		
Db	122	TACTGGGAGGAGGAAATATGCACTCAGAGATGCTTGTACGTTATGTTATGAGGAC 181		

```
QY 309 CCCACTGAGATATGAGGCCCACTCCCGATACGAGCCCCCACTGCGAGATATGG 368
DB 182 CCCACTGAGATATGAGGCCCACTCCCGATACGAGCCCCCACTGCGAGATATGG 241
QY 369 AGCCCAACCCGTAAGAAATGAGGCCCGCTGTGGATACAGAGCTCACTGTGGATA 428
DB 242 AGCCCAACCCGTAAGAAATGAGGCCCGCTGTGGATACAGAGCTCACTGTGGATA 301
QY 429 TGGAGCCCACTCTTGATACGAGGCCCACTGCGAGATATGAGCCCCCACTTACG 488
DB 302 TGGAGCCCACTCTTGATACGAGGCCCACTGCGAGATATGAGCCCCCACTTACG 361
QY 489 ATATGAGCCCC-ACCTCTTGATATGAAACCCCACTCTGCGATATGAGCCCCCACTC 547
DB 362 ATATGAGCCCCCACTCTTGATATGAAACCCCACTCTGCGATATGAGCCCCCACTC 421
QY 548 TGGAGATGAGGCCCACTGCGAGAAATGAGGCCCGCTGGGG 593
DB 422 TGGATATGAGGCCCACTGCGAGAAATGAGGCCCGCTGGGG 467
```

RESULT 2

```
US-09-864-761-31561
; Sequence 31561, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
```

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```
; SEQ ID NO 31561
; LENGTH: 436
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: MAP TO 299716.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
; OTHER INFORMATION: EST HUMAN HIT: AW169980.1, EVALUE 8.00e-11
; OTHER INFORMATION: SWISSPROT HIT: P18616, EVALUE 3.00e-13
US-09-864-761-31561
```

```
Query Match 58.4%; Score 411.4; DB 9; Length 436;
Best Local Similarity 99.8%; Pred. No. 4.3e-125;
Matches 412; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 293 TTATTGTCTATGAGGCCCACTGCGAGATATGAGGCCCACTCTCCGATATCGAGGCC 352
DB 1 TTATTGTCTATGAGGCCCACTGCGAGATATGAGGCCCACTCTCCGATATCGAGGCC 60
QY 353 CACCTGCGAGATATGAGGCCCACTGCGAGAAATGAGGCCCGCTGTGGATACAGAG 412
DB 61 CACCTGCGAGATATGAGGCCCACTGCGAGAAATGAGGCCCGCTGTGGATACAGAG 120
QY 413 CCTCACTGTGTGATATGAGGCCCACTCTTGATATGAGGCCCACTGCGAGATATG 472
DB 121 CCTCACTGTGTGATATGAGGCCCACTCTTGATATGAGGCCCACTGCGAGATATG 180
QY 473 GAGCCCCCACTCTAGATATGAGGCCCACTCTTGATATGAGAAACCCCACTCTGAT 532
DB 181 GAGCCCCCACTCTAGATATGAGGCCCACTCTTGATATGAGAAACCCCACTCTGAT 240
QY 533 ATGAGGCCCACTCTGCGATATGAGGCCCACTGCGAGAAATGAGGCCCGCTGGGG 592
DB 241 ATGAGGCCCACTCTGCGATATGAGGCCCACTGCGAGAAATGAGGCCCGCTGGGG 300
QY 593 GATACAGAGCCCTCACTGCGATATGAGGCCCACTGCGAGAAATGATCAAGAGCCAG 652
DB 301 GATACAGAGCCCTCACTGCGATATGAGGCCCACTGCGAGAAATGATCAAGAGCCAG 360
QY 653 CTCCTGAAAACGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 705
DB 361 CTCCTGAAAACGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 413
```

RESULT 3

```
US-09-764-891-7570
; Sequence 7570, Application US/09764891
; Publication No. US2003007808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7570
; LENGTH: 7099
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7570
```

```
Query Match 53.2%; Score 375.4; DB 11; Length 7099;
Best Local Similarity 99.7%; Pred. No. 1.7e-112;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 290 CAGTTATGCTATATGAGGCCCACTGCGAGATATGAGGCCCACTCTCCGATATCGAG 349
|||||
```


Db 6723 CAGTTATTGCTATGAGAGCCCACTGAGATATGAGCCCACTCCCGATACGAG 6782
Qy 350 CCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAG 409
Db 6783 CCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAG 6842
Qy 410 GAGCTCACTGATGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAT 469
Db 6843 GAGCTCACTGATGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGAT 6902
Qy 470 ATGAGAGCCCACTGATGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTG 529
Db 6903 ATGAGAGCCCACTGATGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTG 6962
Qy 530 GATATGAGAGCCCACTGATGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTG 589
Db 6963 GATATGAGAGCCCACTGATGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTG 7022
Qy 590 CGGATACAGAGCTCTGATGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTG 649
Db 7023 CGGATACAGAGCTCTGATGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTG 7082
Qy 650 AGGCTCTGAAAAACGAG 666
Db 7083 AGGCTCTGAAAAACGAG 7099
RESULT 4
US-09-864-761-15029
; Sequence 15029, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263, 6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15029
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z99716.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
US-09-864-761-15029

Query Match 18.5%; Score 130.4; DB 9; Length 471;
Best Local Similarity 99.2%; Pred. No. 2,8e-32;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 290 CAGTTATTGCTATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGATACGAG 349
Db 340 CAGTTATTGCTATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGATACGAG 359
Qy 350 CCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAG 409
Db 400 CCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAG 459
Qy 410 GAGCTCACTG 421
Db 460 GAGCTCACTG 471

RESULT 5
US-09-764-891-1356
; Sequence 1356, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1356
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1356

Query Match 15.6%; Score 110.2; DB 11; Length 321;
Best Local Similarity 96.8%; Pred. No. 1e-25;
Matches 120; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 1 ATGACATTGATGATGAGAGAACTGATGAGAGAACTGATGAGAGAACTGATGAG 60
Db 195 ATGACATTGATGATGAGAGAACTGATGAGAGAACTGATGAGAGAACTGATGAG 254
Qy 61 ATTAAGGAATATTGAGAGAGCTGATGAGAGAGCTGAGAGAGCTGAGAGAGCTGAG 119
Db 255 ATTAAGGAATATTGAGAGAGCTGATGAGAGAGCTGAGAGAGCTGAGAGAGCTGAG 314

Qy 120 ATTA 123
Db 315 ANTA 318

RESULT 6

US-10-029-386-22839/c
; Sequence 22839, Application US/10029386
; Publication No. US20030194704A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

TITLE OF INVENTION: EXPRESSION ANALYSIS TWO

FILE REFERENCE: AEWICA-X-2

CURRENT APPLICATION NUMBER: US/10/029,386

CURRENT FILING DATE: 2001-12-20

NUMBER OF SEQ ID NOS: 34288

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 22839

LENGTH: 866

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC005630.1

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7

OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 1.9

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4

OTHER INFORMATION: SWISSPROT HIT: Q08379, EVALU 4.00e-10

OTHER INFORMATION: NT HIT: g16161055, EVALU 4.00e-95

OTHER INFORMATION: EST HUMAN HIT: BF828725.1, EVALU 0.00e+00

US-10-029-386-22839

Query Match 15.5%; Score 109; DB 13; Length 866;

Best Local Similarity 54.3%; Pred. No. 4.6e-25; Matches 220; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

Qy 297 TGTCTATGAGGCGCCCACTGACAGATATGAGCCCACTCCCGATATGAGGCCCACTC 356
Db 577 TGTTCACACAGCTC 518
Qy 357 TGCAGATATGAGCCCACTGAGAAATGAGAGCCCGCTGTGGATACAGAGCTTC 416
Db 517 TCTGTTCACAGCTC 458
Qy 417 ACCTGTGCAATGAGAGCCCACTCTGTGATACAGAGCCCACTCTCGATATGAG 476
Db 457 TCTCTCTGTTCACAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 398
Qy 477 CCCACCTCTGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTCGATATGAG 536
Db 397 CT 338
Qy 537 AGCCCACTCTCTGATATGAGAGCCCACTCTGAGAAATGAGAGCCCGCTGTGGATA 596
Db 337 AGCT 278
Qy 597 CAGAGCT 656
Db 277 CAGAGCT 218
Qy 657 TGAAGAGAGCTCA 701
Db 217 TCAATGAGCTCA 173

RESULT 7

US-10-029-386-26033/c

; Sequence 26033, Application US/10029386

; Publication No. US20030194704A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: AEWICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 26033
LENGTH: 1040
TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC005630.1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4

OTHER INFORMATION: NT HIT: g16161055, EVALU 2.00e-97

OTHER INFORMATION: EST HUMAN HIT: BF828725.1, EVALU 0.00e+00

OTHER INFORMATION: SWISSPROT HIT: Q08379, EVALU 6.00e-10

US-10-029-386-26033

Query Match

14.2%; Score 100; DB 13; Length 1040;

Best Local Similarity 53.3%; Pred. No. 4.8e-22;

Matches 211; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

Qy 306 GGGCCCACTCTGAGATATGAGAGCCCACTCTCCGATATGAGAGCCCACTCTGAGATA 365
Db 568 GGCACCTCTCTCTGTTCACAGAGCTCTCTCTGTTCACAGAGCTCTCTCTCTCTCTCA 509
Qy 366 TGAAGCCCACTCTGAGAAATGAGAGCCCGCTGTGGATATGAGAGCTCTCACTGTGCG 425
Db 508 CACAGCTCTCTCTCTGTTCACAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 449
Qy 426 ATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTCT 485
Db 448 TCAAGTATGAGCTCTCTCTCTGTTCACAGAGCTCTCTCTCTCTCTCTCTCTCTCTCT 389
Qy 486 AGATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTCTGATATGAGAGCCCACT 545
Db 388 TGTTCACAGAGCTCTCTCTCTGTTCACAGAGCTCTCTCTCTCTCTCTCTCTCTCTCT 329
Qy 546 TCTGGATATGAGAGCCCACTCTGAGAAATGAGAGCCCGCTGTGGATATGAGAGCTCT 605
Db 328 TCTGTTCACAGAGCTCTCTCTCTGTTCACAGAGCTCTCTCTCTCTCTCTCTCTCTCT 269
Qy 606 ACCTGTGATGAGAGCCCACTCTGAGATATGAGAGCCCACTCTGAGATATGAGAGCC 665
Db 268 TCTCATATGATGAGAGCTCTCTCTCTGTTCATGAGAGCTCTCTCTCTCTCTCTCTCT 209
Qy 666 GGGCT 701
Db 208 CT 173

RESULT 8

US-09-764-891-7571

; Sequence 7571, Application US/09764891

; Publication No. US20030077808A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC006

CURRENT APPLICATION NUMBER: US/09/764,891

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 7571

LENGTH: 291

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-891-7571

QY 306 GGGCCCACTGAGATATGAGACCCCACTCCCGATAGAGAGCCCACTGAGATAT 365
DB 314 GGGAGCTTCTCTGTTTCAACAGAGCTCTCTCTGTTTCAAGAGCTCTCTCTGTTTCA 255
QY 366 TGAAGCCCAACCGTATGAGAAATGAGAGCCCGCTGTGATATGAGAGCTCTCTGTTGCG 425
DB 254 CACAGGCTCTCTCTGTTTCAAGTACAGCTCTCTCTCTGTTTCAACAGAGCTCTCTCTGTT 195
QY 426 ATATGAGAGCCCACTCTGTTGATATGAGAGCCCACTGAGATATGAGAGCCCACTCT 485
DB 194 TCAAGAGCTCTCTCTCTGTTTCAACAGAGCTCTCTCTGTTTCAAGTACAGAGCTCTCTCT 135
QY 486 AGGATATGAGAGCCCACTCTGTTGATATGAGAGCCCACTCTGTTGATATGAGAGCCCACT 545
DB 134 TGTTCACAGAGCTCTCTCTCTGTTTCAAGTACAGAGCTCTCTCTGTTTCAAGTACAGAGCTCT 75
QY 546 TCTCGATATGAGAGCCCACTCTGTTGATATGAGAGCCCACTCTGTTGATATGAGAGCCCTC 605
DB 74 TCTGTTTCAAGTATGAGCTCTCTCTCTGTTTCAAGTACAGAGCTCTCTCTGTTTCAAGAGCTCTC 15
QY 606 ACCTGCTG 613
DB 14 TCTCATG 7

RESULT 11
US-10-029-386-24981
Sequence 24981, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AROMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 24981
LENGTH: 511
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO Z97205.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
US-10-029-386-24981

Query Match 12.5%; Score 87.8; DB 13; Length 511;
Best Local Similarity 52.3%; Pred. No. 3,4e-18;
Matches 194; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
QY 283 CCTTTGCTGTTATTTGTTATGAGAGCCCACTGAGATATGAGAGCCCACTCCCGA 342
DB 138 CTTCTCTCACTGAGAGCTCA 197
QY 343 TACGAGAGCCCACTGAGATATGAGAGCCCAACCGTATGAGAAATGAGAGCCCGCTGTG 402
DB 198 TCCCTCTCTCACTGAGAGCTCA 257
QY 403 GATATGAGAGCT 462
DB 258 GGTCT 317
QY 463 GAGATATGAGAGCCCACTCA 522
DB 318 GAGAGCTCA 377
QY 523 CTTCTGATATGAGAGCCCACTCA 582
DB 378 CTTGAGAGCT 437

QY 583 CCGCTGCGGATATGAGAGCTCA 642
DB 438 CCACCTGAGAGCT 497
QY 643 GAGCCCAAGC 653
DB 498 CTTCTCACTATGC 508

RESULT 12
US-09-864-761-21308/c
Sequence 21308, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aromica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21308
LENGTH: 543
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL009178.4
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
OTHER INFORMATION: EST HUMAN HIT: AM382320.1, EVALU8 8.80e-02
OTHER INFORMATION: SWISSPROT HIT: Q05904, EVALU8 7.00e-12
US-09-864-761-21308

Query Match 12.3%; Score 87; DB 9; Length 543;
Best Local Similarity 54.1%; Pred. No. 6.5e-18;
Matches 177; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 305 GGGCCCCAGCTGACAGATATGAGCCCACTCCGGATAGAGAGCCCACTGACAGAT 364
DB 403 GGGAACTCTCTCCAGTTCCGAGAGACCTTCCAGCTCCGAGAGACTCTCCCACT 344
QY 365 ATGAGCCCAACCCCTAGAAATGAGGCCCGCTGTGGATACAGACTCACTGTGC 424
DB 343 CCGAGAGACTCTCTCCAGTTCCGAGAGACTCTCCAGCTCCGAGAGACTCTCCA 284
QY 425 GATATGAGCCCACTCTTGGATACGAGCCCACTGAGATATGAGCCCACTC 484
DB 283 GTTCCGAGAGACTCTCTCCAGCTCCGAGAGACTCTCCAGTTCCGAGAGACTCTC 224
QY 485 TAGATATGAGCCCACTCTTGGATATGAGCCCACTCTGAGATATGAGCCCACT 544
DB 223 CCACTCCGAGAGACTCTCTCAAACTCTGAGAGACTCTCTCCAGTTCTGAGAGAC 164
QY 545 CTCTCGATATGAGCCCACTCTGAGAGAAATGAGGCCCGCTGCGGATACAGAGCT 604
DB 163 CTCCAGTTCCGAGAGACTCTCTCCAGTTCTGAGAGACTCTCTCCAGTTCCGAGAG 104
QY 605 CACTGCTGATCAGAGCCAGGCTC 631
DB 103 CTCTCCAGTTCCGAGAGACTCTC 77

RESULT 13
US-09-918-995-33536

Sequence 33536, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 33536
LENGTH: 432
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)...(432)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

Query Match 11.0%; Score 77.2; DB 11; Length 432;
Best Local Similarity 62.4%; Pred. No. 9.8e-15;
Matches 121; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 ATGCCATTGATCTGATACCAACTCACTTTGAACAACAGATATTTGCTGCAACTTC 60
DB 231 ATGCCATTATCTCATGAAAGACTGTGATCAAGAGCCCGATTTGGCGAAACTAC 290
QY 61 ATTAAGGAACATATTCGAGAGCTCCATATGTGTGGAGAGCAAGCTACTTTAA 120
DB 291 ATCAAGGAGACATGAGAGCGAGAGGAGGAGGCTGCTGCTCTCAAG 350

QY 121 TTAGCTTCAGAAATGAGATCCATGAATTTGCCAGTTGATGATGAAAGCTGCTCT 180
DB 351 TTGACTTCACGGAGGGGGCCATTGATTTGAGAGCGGAGATGCTCAGGTGCACT 410
QY 181 GCTGTGCCCCGAGG 194
DB 411 CAGCTCTCAGAGG 424

RESULT 14
US-10-029-386-11265

Sequence 11265, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: ABOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11265
LENGTH: 541
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO 297205.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
OTHER INFORMATION: SWISSPROT HIT: P22793, EVALU8 1.00e-04
US-10-029-386-11265

Query Match 9.8%; Score 69.2; DB 13; Length 541;
Best Local Similarity 54.3%; Pred. No. 4.9e-12;
Matches 140; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

QY 308 CCCCACTGAGATATGAGAGCCCACTCCGATACGAGAGCCCACTGAGAGATATG 367
DB 14 CTCACCTGAGAGGCTCTCTCTCTCACTGAGGCTCTCTCTCACTGAGGCTCTT 73
QY 368 GAGCCCAACCGTATGAAATGAGAGCCCGCTGTGGATACAGAGCTCACTGTGAT 427
DB 74 CTCTCCAGCTGAGGCTCTCTCTCTCACTGAGGCTCTCTCTCACTGAGGCT 133
QY 428 ATGAGCCCAACCTCTGATACGAGAGCCCACTGAGATATGAGAGCCCACTGTAG 487
DB 134 CTCTCTTCCAGCTGAGGCTCTCTCTCTCACTGAGGCTCTCTCTCACTGAG 193
QY 488 GATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCCCACTC 547
DB 194 GCTCCCTCTCTCACTGAGGCTCTCTCTCTCACTGAGGCTCTCTCTCTCAATG 253
QY 548 TCGATATGAGAGCCCACT 565
DB 254 CAGGCTCTCTGAGAGCC 271

RESULT 15

US-10-029-386-24981/c
Sequence 24981, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: ABOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20

```
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 24981
; LENGTH: 511
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO 297205.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
US-10-029-386-24981
```

```
Query Match          9.6%; Score 67.8; DB 13; Length 511;
Best Local Similarity 50.8%; Pred. No. 1.4e-11;
Matches 162; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
```

```
QY      313  CCTGCAGATATGAGAGCCCACTCCCGATACGAGCCCACTGCAGAGATATGAGCC 372
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      343  CCTGCAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 284
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      373  CAACCCGTAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 432
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      283  GAGCCTGCAAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 224
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      433  GCCCACTCTTGGATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 492
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      223  GGGAGAGCTTGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 164
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      493  GAGAGCCCACTCTTGGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 552
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      163  GAGAGGAGAGCTTGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 104
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      553  TATGAGAGCCCACTTGCAGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 612
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      103  GAGAGAGGAGAGCTTGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAT 44
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY      613  GATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 631
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      43  TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 25
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: December 16, 2003, 17:58:46
Job time : 233.016 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:21:45 ; Search time 1450.89 Seconds
(without alignments)
11809.735 Million cell updates/sec

Title: US-09-864-291-11_COPY_1_705

Perfect score: 705

Sequence: 1 atgcatttcattcgtcgtac.....cctcttcacgtccatttc 705

Scoring table: IDENTITY_NUC

Gapex 10.0, Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_estbm:*
3: em_estbn:*
4: em_estbu:*
5: em_estbv:*
6: em_estpl:*
7: em_estro:*
8: em_hnc:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_hnc:*
12: gb_estc3:*
13: gb_estc4:*
14: gb_estc5:*
15: em_estfm:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vtc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rnd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	656.4	93.1	1079	12	BMS64167
2	368.2	52.2	663	10	BG693398
3	362.8	51.5	775	10	BG722815
4	333.2	47.3	1414	11	AK015863

Result No.	Score	Query Match	Length	ID	Description
5	263	37.3	701	10	BG701881
6	252.2	35.8	717	14	BY715414
7	169.8	24.1	814	13	BU961805
8	160.4	22.8	488	28	AQ997933
9	138.8	19.7	426	28	AQ997939
10	117.2	16.6	671	29	AG062842
11	104.2	14.8	464	10	BF000754
12	101	14.3	488	10	BF476944
13	100.6	14.3	568	12	BI390687
14	100.6	14.3	636	12	BM490729
15	98	13.9	499	13	BQ356763
16	97.4	13.8	560	13	BU103740
17	97.4	13.8	700	13	BU284977
18	97.4	13.8	723	9	AJ452591
19	97.4	13.8	744	13	BU230211
20	97.4	13.8	932	13	BU107273
21	90.4	12.8	706	13	BU433499
22	89.8	12.7	773	9	AV899050
23	88.8	12.6	755	13	BM089605
24	88.4	12.5	772	9	AJ454950
25	88	12.5	611	9	AL846063
26	87.6	12.4	686	12	BP021598
27	87.2	12.4	594	12	BM190972
28	87.2	12.4	629	10	BF828725
29	87.2	12.4	816	13	BU916890
30	87.2	12.4	931	13	BU914288
31	86.8	12.3	562	12	BM495739
32	86.6	12.3	622	10	BE468886
33	86.6	12.3	899	10	BE470466
34	86.4	12.3	613	9	AL849211
35	86.4	12.3	635	9	AL889954
36	86.4	12.3	647	9	AL849651
37	86.4	12.3	655	9	AL898723
38	86.4	12.3	661	9	AL972453
39	86.4	12.3	663	9	AL881810
40	86.4	12.3	663	9	AL896682
41	86.4	12.3	735	12	BP019748
42	86.4	12.3	823	13	BM087191
43	85.6	12.1	460	10	BE680329
44	85.6	12.1	509	10	BG515416
45	85.6	12.1	607	12	BM190975

ALIGNMENTS

RESULT 1
BMS64167
LOCUS
DEFINITION
AGENCOURT_6560149 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742129
5', mRNA sequence.
BMS64167
VERSION
BMS64167.1 GI:18611738
KEYWORDS
SOURCE
EST.
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1079)
NIH-MGC http://mgc.nci.nih.gov/
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
CONTACT: Robert Strausberg, Ph.D.
COMMENT
Email: cgapsb-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
plate: LHAM12759 row: m column: 10
High quality sequence stop: 726.

FEATURES
source

Location/Qualifiers
1. 1079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5742129"
/issue_type="medulla"
/lab_host="DH10B"
/clone_lib="NIH_MGC_119"
/note="Organ: Brain; Vector: pCMV-Sport6; Site: 1; Nct1; Site: 2; Score (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (BcoRI site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH MGC Library."

BASE COUNT 267 a 297 c 268 g 247 t
ORIGIN

Query Match 93.1%; Score 656.4; DB 12; Length 1079;
Best Local Similarity 97.3%; Pred. No. 1.8e-170;
Matches 689; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

1 ATGCCATTGATCTGATGAGAACTCCTGTTGAACAACAGTATTGCTGCAACTTC 60
269 ATGCCATTGATCTGATGAGAACTCCTGTTGAACAACAGTATTGCTGCAACTTC 328
61 ATTAAGGAAGCTATTCAGGAGCTCCATATGATGCTGGGAGAGACAGTACTTTTAA 120
329 ATTAAGGAAGCTATTCAGGAGCTCCATATGATGCTGGGAGAGACAGTACTTTTAA 388
121 TAGTCTTCAGAAATGAGATGCGATTGATTTGCCAGTTGATGATGAGAGCTGCTCT 180
389 TAGTCTTCAGAAATGAGATGCGATTGATTTGCCAGTTGATGATGAGAGCTGCTCT 448
181 GCTGTGCCCCGAGAGATTTCCATTGAACTTAAATGATGCTGCTGATGAGAAAT 240
449 GCTGTGCCCCGAGAGATTTCCATTGAACTTAAATGATGCTGCTGATGAGAAAT 508
241 TATGTAATTAATCTGGGAGAGAAATATGTCACCTCCAGATGCGATGATGATGTC 300
509 TATGTAATTAATCTGGGAGAGAAATATGTCACCTCCAGATGCGATGATGATGTC 568
301 TATGAGGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCACTGCA 360
569 TATGAGGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCACTGCA 628
361 GATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCACTGCA 420
629 GATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCACTGCA 688
421 GTGCGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAG--CCCC 479
689 GTGCGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCACT 748
480 ACCTCTAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGC 539
749 ACCTCTAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGC 808
540 CCCAGCTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAG 599
809 CCCAGCTCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAG 868
600 AGCTCACTGCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAG 657
869 AGCTCACTGCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAG 928
658 GAAAGAGAGCT 705
929 GAAAGAGAGCT 976

RESULT 2
BG699398 663 bp mRNA linear EST 07-MAY-2001
602679081P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5',
DEFINITION
LOCUS
ACCESSION
BG699398
VERSION
BG699398.1 GI:13967653
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 663)
NIH-MGC <http://mgs.mci.nih.gov/>.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLES
Unpublished
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Tohshiyuki and Piero Carninci (RIKEN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLNL10703 row: a column: 21
High quality sequence stop: 660.

FEATURES
sourceLocation/Qualifiers
1. 663

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4811804"
/issue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtagag
); Oligo-dT primed using primer 5'-TTTCTTTTCTTTTCTT-3',
size-selected for average insert size 2.5 kb and
normalized to RQF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 168 a 158 c 169 g 164 t 4 others
ORIGIN

Query Match 52.2%; Score 368.2; DB 10; Length 663;
Best Local Similarity 97.1%; Pred. No. 7.6e-91;
Matches 395; Conservative 0; Mismatches 10; Indels 2; Gaps 2;

1 ATGCCATTGATCTGATGAGAACTCCTGTTGAACAACAGTATTGCTGCAACTTC 60
269 ATGCCATTGATCTGATGAGAACTCCTGTTGAACAACAGTATTGCTGCAACTTC 317
258 ATGCCATTGATCTGATGAGAACTCCTGTTGAACAACAGTATTGCTGCAACTTC 317
61 ATTAAGGAAGCTATTCAGGAGCTCCATATGATGCTGGGAGAGACAGTACTTTTAA 120
318 ATTAAGGAAGCTATTCAGGAGCTCCATATGATGCTGGGAGAGACAGTACTTTTAA 377
121 TAGTCTTCAGAAATGAGATGCGATTGATTTGCCAGTTGATGATGAGAGCTGCTCT 180
378 TAGTCTTCAGAAATGAGATGCGATTGATTTGCCAGTTGATGATGAGAGCTGCTCT 437
181 GCTGTGCCCCGAGAGATTTCCATTGAACTTAAATGATGCTGCTGATGAGAAAT 240
438 GCTGTGCCCCGAGAGATTTCCATTGAACTTAAATGATGCTGCTGATGAGAAAT 497
241 TATGTAATTAATCTGGGAGAGAAATATGTCACCTCCAGATGCGATGATGATGTC 300
498 TATGTAATTAATCTGGGAGAGAAATATGTCACCTCCAGATGCGATGATGATGTC 557

Oy 301 TATGGGGCCCCACCTGAGATATAGAGACCCCACTCCCGATAGAGAGCCCACTGCA 360
 Db 558 TATGAGGCCCACTGAGATATAGAGACCCCA-CTCCCGATAGAGAGCCCA-CTGCA 615
 Oy 361 GGATATAGAGCCCAACCCGTAGAAATAGAGCCCGCTGTGGATA 407
 Db 616 GGATATAGAGCCCAACCCGTAGAAATAGAGCCCGCTGTGGATA 662
 RESULT 3
 Bg722815 775 bp mRNA linear EST 08-MAY-2001
 LOCUS Bg722815
 DEFINITION 602695278P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5',
 mRNA sequence.
 ACCESSION Bg722815
 VERSION Bg722815.1 GI:14002002
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 775)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgsabp-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shizaki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.llnl.gov
 Plate: LHAM10743 row: 1 column: 20
 High quality sequence stop: 651.
 Location/Qualifiers
 1..775
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4827427"
 /lab_host="DH10B"
 /clone_id="NIH_MGC_97"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
 size-selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 191 a 212 c 193 g 179 t
 ORIGIN
 Query Match 51.5%; Score 362.8; DB 10; Length 775;
 Best Local Similarity 95.1%; Pred. No.2.6e-89;
 Matches 428; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

Oy 181 GCTGTGCCCGAGGATTTCACTTAGAACCTTAA--TGACTGTTCAGCTCTATAGGAAT 239
 Db 456 GCTGTGCCCGAGGATTTCACTTAGAACCTTAACTGATCTGTGCTCTATAGGAAT 515
 Oy 240 TTAGTAATTAATCTGGGAGAGGAATATGTGACTTCCAGATGCTTTGTCACTTATGT 299
 Db 516 TTAGTAATTAATCTGGGAGAGGAATATGTGACTTCCAGATGCTTTGTCACTTATGT 575
 Oy 300 CTATGGGGCCCCACCTGAGAG-ATATGAGACCCCACTCCCGATATAGAGAGCCCACTG 358
 Db 576 CTATGAGAGCCCACTGAGAGCATATAGAGACCCCACTCCCGATATAGAGAG-CCCACTG 634
 Oy 359 CAGGATATAGAGCCCAACCCGTAGAAATAGAGAGCCCGCTGTGGATATAGAGCCCTC-A 417
 Db 635 CAGGATATAGAGCCCAACCCGTAGAAATAGAGAGCCCGCTGTGGATATAGAGCCCTCAA 694
 Oy 418 CTTGTGCGA-TATGAGACCCCACTCTTGG 446
 Db 695 CTTGTGCGAATAGAGAGCCCACTCTCGG 724
 RESULT 4
 AK015863
 LOCUS AK015863
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
 library, clone:4930521i23 product:hypothetical protein, full insert
 sequence.
 ACCESSION AK015863
 VERSION AK015863.1 GI:12854363
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Kikunishi, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsunoto, H., Sakauechi, S., Ikegami, T., Kashiwagi, K.,
 Fujisawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, D.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 4
 Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
 Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,
 Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakata, I.,
 Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,
 Kadota, K., Matsuoka, H., Ashburner, M., Batalov, S., Casavant, T.,
 Fleischmann, W., Gaasterland, T., Glisic, C., King, B., Kochiwa, H.,
 Kuehl, P., Lewis, S., Matsuo, Y., Nishida, I., Pesole, G.,
 Quackenbush, J., Schriml, L.M., Staudt, P., Suzuki, R., Tomita, M.,
 Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
 Baldarelli, R., Barsh, G., Blake, J., Botelli, D., Bojunga, N.,

Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,
Pletcher, C., Fujita, K., Garbold, M., Gustinchik, S., Hill, D.,
Hofmann, M., Hunne, D. A., Kaniyva, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashim, J., Mazzarelli, J., Mombere, P.,
Ring, B., Ringwald, N., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schobach, C., Seya, T., Shiota, Y., Storch, K. P., Suzuki, H.,
Toyooka, K., Wang, K. H., Wetz, C., Whiteaker, C., Whiting, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, I., Kawaji, H., Kontseki, S.,
and Hayashizaki, Y.

TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)

MEDLINE 21085660
PUBMED 11217851

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens	1 (bases 1 to 701)	NIH-MGC	http://mgc.nci.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished
Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
NIH-MGC	http://mgc.nci.nih.gov/				
Unpublished					
Contact: Robert Strausberg, Ph.D.					
Email: cgsab@remail.nih.gov					
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.					
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihaki Toshiyuki and Piero Carninci (RIKEN)					
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)					
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:					
http://image.lnl.gov					
Plate: LLM10713 row: j column: 10					
High quality sequence atop: 660.					
Location/Qualifiers					
1..701					
/organism="Homo sapiens"					
/mol_type="mRNA"					
/db_xref="taxon:9606"					
/clone="IMAGE:4815843"					
/cissue_type="hippocampus"					
/lab_host="DH10B"					
/clone_id="NIH_MGC_95"					
/note="Organ: brain; Vector: pBluescriptpr (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (grecag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size selected for average insert size 2.5 kb and normalized to 10^5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."					
BASE COUNT	176 a	164 c	189 g	172 t	
ORIGIN					
Query Match	37.3%	Score 263;	DB 10;	Length 701;	
Best Local Similarity	89.4%	Pred. No. 1e-61;			
Matches 387;	Conservative 0;	Mismatches 30;	Indels 16;	Gaps 9	
1 ATGCCATTGATCTGATGACGAACCTC--ACTGTTGAACAACGATATTGCT--GCAA 55					
261 ATGCATTTGATCTGATGACGAATCTTAACTGTTGAACAACGATATTGCTGCGAAA 320					
56 ACTTCATTAAAGGAACATATTCAGGACGCTCATATGCTGCTGGAAAGCAAGCTACTT 115					
321 CTTCGATTAAAGGAACATATTCAGGACGCTCCATATGCTGCTGGAAAGCAAGCTACT 380					
116 TT----AAATTAGTCTCAGAAATGGAATGCAATGATTTGCCAGTGAATGGTAA 171					
381 TTATTAATATGTAATCTTCAAAATGGAATGCAATGATTTGCCAGTGAATGGTAA 440					
172 GCTGCTCTGCTGCTGCTGCCAGATTTCCATTTAGAACCTTAA--ATGACTGTTGAGCTC 230					
441 GCGGCTCTGCTGCTGCTGCCAGATTTCCATTTAGAACCTTAAATGATGATGCTGCTC 500					
231 TATGGAAATTTATGTAATTTACTGGGAAGGAATATGCA--CTCCACAGATGCTTGT 288					
501 TATGGGCAATTTATGTAATTTACTGGGAATGGAATTTGCAATGCTCCACAGATGCTTGT 560					
289 TCAGTATTTGCTAATGGGGCCCACTGCAAGATATGAGAGCCCACTCCGGATATGCGA 348					
561 AAGTTA--TGCTATGAGAGCGCACTGGGGAATGAGAGCCCACT--CCGATATGCGA 618					
349 GCCCACTGAGAGATATGAGAGCCCACTCGTAGGAAATGAGAGCCGCTGTGGATATC 408					
619 GCCCA--CTGCAAGATATGAGAGCCCACTCGTAGGAAATGAGAGCCG--CTGTGGATATC 676					
409 AAGCTTCACTG 421					

Db	717 bp	mRNA	linear	EST
RESULT 6				BY715414
LOCUS	BY715414			
DEFINITION	BY715414 RIKEN full-length enriched, adult male testis Mus musculus			
ACCESSION	BY715414			
KEYWORDS	CDNA clone 4930521123 5', mRNA sequence.			
SOURCE	BY715414.1 GI:27128531			
ORGANISM	Mus musculus (house mouse)			
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 717) Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Oseato,N., Saito,K., Suzuki,H., Yamataka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bulc,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chochoja,C., Coban L.B., Cousins,S., Dalla,E., Dreagan,T.A., Flecher,C.F., Forrest .A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A. Gough,J., Grimmond,S., Guetincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kamai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagasa,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Mik H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Perlea,G., Peocle,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou M., Shimada,M., Sultana,R., Takemaki,Y., Taylor,M.S., Teasdale R.D., Tomita,M., Verdaro,R., Wagner,T., Wahlstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wyshaw-Boris,A., Yanagisawa M., Yang,I., Yang,L., Yuan,Z., Zaylan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirotsane-Kishikawa,T., Komno,H., Nakamura M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Saeki,D., Shibba K., Shingawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander R.S., Rogers,J., Birney,E. and Hayashizaki,Y. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 22354683 12466851			
TITLE	Contact: Yoshihide Hayashizaki			
JOURNAL	Laboratory for Genome Exploration Research Group, RIKEN Genomic			
MEDLINE	Sciences Center (GSC), Yokohama Institute			
PUBMED	The Institute of Physical and Chemical Research (RIKEN)			
COMMENT	1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/ Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda S., Hashizume,W., Hayashida,K., Hirozane,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Komno H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,D., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Saeki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watabiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1751-1771 (2000)			

Db	243	CCTCTGATGATCAAGGGCCACGTCGTGTAGACACAGGAAGTCCACCTGCTGCGGTCTGAAGCA	302
QY	625	AGCGCTCAGGAATCTACACAGCGCCGAGCTCTGTAACAGAGCTTCTCTTCCCTCTGCC	684
Db	303	GGCATTCCCAATGCTGTGGCAGTCCAGAACTCTGAATTCAGAGATCTTTTCCCTTAC	362
QY	685	TCCTCTCTCAGGTCCATCTT	705
Db	363	TCGTCTCAAGATCCACTCT	383

RESULT 8			
LOCUS	AC097933/c	488 bp	DNA linear GSS 24-FEB-2000
DEFINITION	RPCT-23-384A20.TU RPCT-23 Mus musculus genomic clone RPCT-23-384A20		
ACCESSION	AC097933		
KEYWORDS	'genomic survey sequence.'		
SOURCE	AC097933.1 GI:7072953		
ORGANISM	Mus musculus (house mouse)		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 488)		
TITLE	Zhao,S., Nieman,W., Paldbjyrm,T., Malek,U., Shatman,S., Akinet		
JOURNAL	,B., Levins,M., McGinn,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.		
COMMENT	Mouse BAC End Sequences from Library RPCT-23		
	Unpublished		
	Other GSSs: RPCT-23-384A20.TU		
	Contact: Shaying Zhao		
	Department of Eukaryotic Genomics		
	The Institute for Genomic Research		
	9712 Medical Center Dr., Rockville, MD 20850, USA		
	Tel: 301 838 0200		
	Fax: 301 838 0208		
	Email: szhao@igr.org		
	Clones are derived from the mouse BAC library RPCT-23. For BAC		
	library availability, please contact Pieter de Jong		
	(pieter@djong.med.buffalo.edu). Clones may be purchased from		
	BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)		
	or from Resea ch Genetics (info@resgen.com). BAC end page:		
	http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html		
	Plate: 384 row: A column: 20		
	Seq primer: SP6		
	Class: BAC ends.		

FEATURES			
source	location/Qualifiers		
	1..488		
	/organism="Mus musculus"		
	/mol_type="genomic DNA"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="RPCT-23-384A20"		
	/sex="Female"		
	/lab_host="DH10B"		
	/clone_1lb="RPCT-23"		
	/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1:		
	Ecotri; Site 2: Ecotri; Female C57BL/6J mouse kidney and/or		
	brain genomic DNA was isolated and partially digested		
	with a combination of Ecotri and Ecotri Methylase. The		
	selected DNA was cloned into the pBAC3.6 vector at the		
	DH10B sites. The ligation products were transformed into		
	DH10B electrocompetent cells (BRL life Technologies). "		

BASE COUNT	102 a	128 c	157 g	100 t	1 others
ORIGIN					
Query Match	22.8%	Score 160.4;	DB 28;	Length 488;	
Best Local Similarity	67.3%;	Pred. No. 2e-33;			
Matches 243;	Conservative	0;	Mismatches 112;	Indels 6;	Gaps 1;

QY	289	TCAGTATTATCTCTAAGGGGCCCCACCTGACAGATATGAGACCCCACTCCCGATATACGA	348
Db	457	TCACCCGGAGATACGAGATCCCATCTGACGGATATGAGAGCCCACTGCGGGATACGA	398

FEATURES	SOURCE
LOCUS	AC097939/c
DEFINITION	AC097939 426 bp DNA linear GSS 24-FEB-2000 RPCI-23-384A21.TU RPCI-23 Mus musculus genomic clone RPCI-23-384A21
ACCESSION	AC097939
VERSION	AC097939.1 GI:7072959
KEYWORDS	GSS.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus. 1 (bases 1 to 426) Zhao S., Nieman, W., Paldblyum, T., Malek, J., Shatman, S., Akintet , B., Levins, M., McGann, S., Tesgaye, G., Geer, K., Krol, M., de Jong, P. and Frazer, C.M. Mouse BAC End Sequences from Library RPCI-23 Unpublished Other GSSs: RPCI-23-384A21.TV Contact: Shaying Zhao Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: szhao@tigr.org
TITLE	Clones are derived from the mouse BMC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://BACPAC.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac ends/mouse/bac_end_intro.html Plate: 364 row: A column: 21 Seq primer: SP6 Class: BAC ends.
JOURNAL	Location/Qualifiers
COMMENT	1. .426 /organism="Mus musculus" /mol_type="genomic DNA" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="RPCI-23-384A21" /sex="Female" /lab_host="DH10B" /clone_1ib="RPCI-23" /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRII MethyIase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

Query Match 19.7%;	Score 138.8;	DB 28;	Length 426;
Best Local Similarity 66.9%;	Pred. No. 1.7e-27;		
Matches 230;	Conservative 0;	Mismatches 107;	Indels 7;
			Gaps 2;

Qy	307	GCCCCACCTGAGGATATGAGAGCCCACTCCGGATACGAGAGCCCACTGCAGAAAT	366
Db	425	GTCCATCTGACGGGTATGAGAGCCCACTGGGGGTACGAGAGCCCGCCAGTGGTAT	366
Qy	367	GGAGCCCCAACCCGTAGAAATGAAAGCCCGCTGTGGATACAGAGCTCACTTGGCA	426
Db	365	GTAGCCCATCTCCGGATACGATCTCTGGCTCTGTGATACGGAG-----CTGTAGA	312
Qy	427	TATGAGGCCCACTCTTGATATACGAGAGCCCACTGCAGATATGAGAGCCCACTCTA	486
Db	311	TATGGCTCCCTCCTCTATATGTAGCTACCCCAATGGGGTATGAGAGATCACTCT	252
Qy	487	GGATATGAGAGCCCACTCTTGATATGGAACCCCACTCTTCGATATGAGAGCCCACT	546
Db	251	GGATATGAGAGCCCACTGTGAGATACGATATCCCACTCTCGATATGAGAGCCCAAT	192
Qy	547	CTCGGATATGAGAGCCCACT- <u>GCAGGAAATGAAAGCCCGCTGGGATACAGAGCTC</u>	605
Db	191	ATGAGATGAGGAGAAACAATCTCCATGATACGGAACATACTATGGATCCGATCCCC	132
Qy	606	ACCTGCTGATCAGAGAGCAGGCTCTAGAAATCTACAGAGCCC	649
Db	131	ACCTCCGATATGAGAGCCCGCATATGGGGATGAAATCTCTC	88

AG062842	LOCUS	DEFINITION	ACCESSION	VERSION
AG062842	671 bp	DNA	1 linear	GSS 03-NOV-2001
AG062842	Pan troglodytes DNA, clone: PTB-051P01.R, genomic survey sequence.			
AG062842.1	GI:1614644			

SOURCE ORGANISM	Pan troglodytes (chimpanzee)
REFERENCE AUTHORS	1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	BAC end sequences of library PTB
JOURNAL	Unpublished
REFERENCE AUTHORS	2 (bases 1 to 671) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Shuhiro-chou, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chlimbse@sc.riken.go.jp, URL:htp://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

```

/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-051P01.R"
/sex="male"
/cell_type="lymphoblast"
/clone_id="PTB Chimpanzee Male BAC library"
BASS COUNT 77 a 270 c 103 g 220 t 1 others
ORIGIN

```

Query Match	16.6%	Score 117.2	DB 29	Length 671
Best Local Similarity	55.6%	Pred. No. 2.1e-21		
Matches 224	Conservative 0	Mismatches 179	Indels 0	Gaps 0

Oy	297	TGTCATAGAGGCCCACTGAGAGATATGAGGCCCACTCCCGGATATAGAGGCCCAAC	356
Dd	196	TGTTACAGTAGCCTCTCCTCTGTTCAATGAGCCTCTCCTCTGTTACAGTAGCCTCTCC	255
Oy	357	TGCAGATATAGAGCCCAACCGTAGAATAAGAGCCCGCTGAGATACAGAGCTTC	416
Dd	256	TTCTGTTACAGTAGCCTCTCCTCTGTTCAATGAGCCTCTCCTCTGTTACAGTAGCCTC	315
Oy	417	ACCTGTGAGATATGAGGCCCACTCTTGGATAGAGGCCCACTGACAGATATAGAGC	476
Dd	316	TCTCTCTGTTACAGTAGCCTCTCCTCTGTTACAGTAGCCTCTCTCTCTGTTCAATGAGC	375
Oy	477	CCCACTCTAGATATGAGGCCCAACCTTGGATATAGAAACCCCAACCTCTGAGATATAG	536
Dd	376	CTCTCTCTCTTCAATGAGCCTCTCCTCTGTTCAATGAGCCTCTCTCTTCTGTTCAATG	435
Oy	537	AGCCCACTCTCGAGATATGAGGCCCACTGACAGAAATAGAGCCCGCTGACGAGATA	596
Dd	436	AGCCCTCTCTCTGTTCAATAGGCTCTCCTCTGTTCAATGAGCCTCTCCTCTGTTCA	495
Oy	597	CAGAGCTTCACTGTGATCAGAGACGAGCCTCAGAAATCAAGCAGCCAGGCTCC	656
Dd	496	CACAGCCTCTCTCTGTTCAATGAGCCTCTGCTCTCTTCAAGTAGCCTCTCTTCTGT	555
Oy	657	TGAAAGAGGCTTCTCTTCCCTGCTGCTCCCTCTCTTCAAGC	699
Dd	556	TCACGTAGCCTCTCTCTCTGTTCAAGTAGCCTCTCTCTCTGTC	598

RESULT 11	
BP000754	
LOCUS	464 bp mRNA linear EST 06-OCT-2001
DEFINITION	7143901.x1 NC1-CGAP B16 Homo sapiens cDNA clone IMAGE:3337488 3' similar to contains element YARI repetitive element ;, mRNA sequence.
ACCESSION	BP000754
VERSION	BP000754.1 GI:10701029
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 464)
AUTHORS	NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, ph.D.

FEATURES	Location/Qualifiers
Source	1. .671

FEATURES
source

Location/Qualifiers
1..464
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3337488"
/sex="female"
/tissue_type="lobular carcinoma in situ"
/dev_stage="adult"
/lab_host="DH10B"
/clone_1lb="NCI CGAP Br16"
/note="Organ: breast; Vector: pAMP1; mRNA made from breast carcinoma tissue, cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 400 bp. Primary library, non-amplified."

BASE COUNT 58 a 202 c 59 g 145 t
ORIGIN

Query Match 14.8%; Score 104.2; DB 10; Length 464;
Best Local Similarity 53.9%; Pred. No. 6.8e-18;
Matches 214; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY 306 GAGCCCACTGACGATATGAGGCCCACTCCGATACGAGCCCACTGACGATTA 365
DB 58 GCGAGCTTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTCA 117
QY 366 TGGAGCCCACTGACGATATGAGGCCCACTGTTGGAATACAGAGCTTCACTTTCG 425
DB 118 CACAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTCTCT 177
QY 426 ATATGAGGCCCACTGTTGATACGAGGCCCACTGACGATATGAGGCCCACTCT 485
DB 178 TCAGTACCTCTCTCTCTGTTTACAGAGCTCTCTCTATATGTTTACAGAGCTCTCTCC 237
QY 486 AGATATGAGGCCCACTCTTGTATGAGAACCCCACTCTCGATATGAGGCCCACT 545
DB 238 TGTTCACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCC 297
QY 546 TCTCGATATGAGGCCCACTGACGATATGAGGCCCACTGTTGGAATACAGAGCTCTC 605
DB 298 TCTGTTTACAGAGCTCTCTCTCTGTTTACAGAGCTCTCTCTCTGTTTACAGAGCTCTC 357
QY 606 ACCGCTGATACAGAGGCCCACTGACGATATGAGGCCCACTGTTGGAATACAGAGCTCT 665
DB 358 TCTCATATGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCT 417
QY 666 GAGCTTCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCTGTTTCA 702
DB 418 CTCCTCTCTCTCTCTGTTTACAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 454

RESULT 12
BF476944/c 488 bp mRNA linear EST 05-DEC-2000
LOCUS na556d10.x1 NCI CGAP Br16 Homo sapiens cDNA clone IMAGE:3260227 3'
DEFINITION similar to contains element YARI repetitive element ; mRNA
sequence.

ACCESSION BF476944
VERSION BF476944.1 GI:11547771
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapb-remail.nih.gov
Tissue Procurement: Dennis Sgroi, M.D., Kristina Cole, M.D., Ph.D.

FEATURES
source

student, Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lemon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E.B. Consortium/LINT, send email to:
info@image.jhmi.gov
Seq primer: -40RP from Glbco
High quality sequence stop: 440.

Location/Qualifiers
1..488
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3260227"
/sex="female"
/tissue_type="lobular carcinoma in situ"
/dev_stage="adult"
/lab_host="DH10B"
/clone_1lb="NCI CGAP Br16"
/note="Organ: breast; Vector: pAMP1; mRNA made from breast carcinoma tissue, cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 400 bp. Primary library, non-amplified."

BASE COUNT 154 a 68 c 204 g 62 t
ORIGIN

Query Match 14.3%; Score 101; DB 10; Length 488;
Best Local Similarity 53.1%; Pred. No. 5.4e-17;
Matches 215; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 297 TGTCTATGAGGCCCACTGACGATATGAGGCCCACTCCGATACGAGGCCCACT 356
DB 479 TGTTCACAGAGCTCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCTCC 420
QY 357 TGGAGATATGAGGCCCACTGACGATATGAGGCCCACTGTTGGAATACAGAGCTCTC 416
DB 419 TCTGTTTACAGAGCTCTCTCTCTGTTTACAGAGCTCTCTCTATATGTTTACAGAGCTCTC 360
QY 417 ACTGTCGATATGAGGCCCACTCTTGTATGAGAACCCCACTCTCGATATGAGGCCCACT 476
DB 359 TCTCTGTTTACAGAGCTCTCTCTCTGTTTACAGAGCTCTCTCTGTTTACAGAGCTCTCT 300
QY 477 CCACTCTGATATGAGGCCCACTCTTGTATGAGAACCCCACTCTCTGATATGAGGCCCACT 536
DB 299 CTCCTCTCTGTTTACAGAGCTCTCTCTCTGTTTACAGAGCTCTCTCTCTCTCTCTCTCT 240
QY 537 AGCCCACTCTGATATGAGGCCCACTGACGATATGAGGCCCACTGTTGGAATACAGAGCTCT 596
DB 239 AGCT 180
QY 597 CAGAGCT 656
DB 179 CAGAGCT 120
QY 657 TGAAGAGAGCTTCT 701
DB 119 TCAGTGTCT 75

RESULT 13
B1390687 568 bp mRNA linear EST 06-AUG-2001
LOCUS ppbin.pk001.p22 Normalized chicken pituitary/Hypothalamus/pinal
DEFINITION library Gallus gallus cDNA clone ppbin.pk001.p22 5' similar to
gb|U09511.1 (U09458) WW domain binding protein-2 [Homo sapiens].

ACCESSION B1390687
VERSION B1390687.1 GI:15083969
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus

RESULT 14	
BM490729	
LOCUS	636 bp mRNA linear EST 07-FEB-2002
DEFINITION	ppgzn.pk004.e6 Normalized Chicken Pituitary/Hypothalamus/Pineal Library (ppgzn) Gallus gallus cDNA clone ppgzn.pk004.e6 5' similar to gb AB010951.1 (U79458) WW domain binding protein-2 [homo sapiens], mRNA sequence.
ACCESSION	BM490729
VERSION	BM490729.1 GI:18611660
KEYWORDS	EST.
SOURCE	Gallus gallus (chicken)
ORGANISM	Gallus gallus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE	1 (bases 1 to 636)
AUTHORS	Potter,T.B. and Cogburn,L.A.
TITLE	ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA

RESULT	15
BQ356763	
LOCUS	BQ356763 499 bp mRNA linear EST 20-MAY-2002
DEFINITION	I13-HT0618-120500-138-A10 HT0618 Homo sapiens cDNA, mRNA sequence.
ACCESSION	BQ356763
VERSION	BQ356763.1 GI:21022003
KEYWORDS	RST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 499) Nagai,M.A., de Garcia Correa,R., Veitovskí-Almeida,S., Britones,M.R., Das Neto,E., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.P., Matsukuma,A., Baita,G.S., Simpson,D.H., Brumbeil,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare, 'M.J.', Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
PUBMED	10737800
COMMENT	Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the PAPSP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&cl=IL3-HT0618-120500-138-A10&3=2000-05-12&t=1>)

Seq primer: puc 18 forward

High quality sequence start: 15

High quality sequence stop: 162.

Location/Qualifiers

1..499

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="HT0618"

/note="Organ: head, neck; Vector: puc18, Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 62 a 216 c 66 g 155 t

ORIGIN

Query Match 13.9%; Score 98; DB 13; Length 499;

Best Local Similarity 54.5%; Pred. No. 3.7e-15;

Matches 217; Conservative 0; Mismatches 180; Indels 1; Gaps 1;

```
QY 306 GGGCCCACTGCAAGATATGAGAGCCCACTCCGATAGAGAGCCCACTGAGAGATA 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 GGAAGCTTCTCTGTTCAACAGAGCTCTCTCTGTTCAAGTAGAGCTCTCTGTTCA 103
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 366 TGAAGCCAAACCGTAGAAATGAGAGCCGCTGTGGATACAGAGCTTCACT-GTGC 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 CACAGCTCTCTCTGTTCAAGAGCTCTCTCTGTTCAACAGAGCTCTCTCTCTGT 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 425 GATATGAGAGCCCACTCTGATAGAGAGCCCACTGAGAGATATGAGAGCCCACTC 484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 164 TATACATAGCTCTCTCTGTTCAACAGAGCTCTCTCTCAATTCAGTAGAGCTCTCT 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 485 TAGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTGATATGAGAGCCCACTC 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 224 CTGTTCAACAGAGCTCTCTCTGTTCAAGTAGAGCTCTCTCTGTTCAATGAGCTCTC 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 545 CTCTCGAATATGAGAGCCCACTGAGAGAAATGAGAGCCGCTGTGGATACAGAGCT 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 284 CTCTGTTCAAGAGCTCTCTCTGTTCAACATAGAGCTCTCTCTGTTCAACAGAGCT 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 605 CACCTGTGATACAGAGCCAGAGCTCAGAGATTCACAGAGCCAGAGCTCTGAAACG 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 CTCTCAATGTTCAAGTAGAGCTCTCTCTGTTCAACAGAGCTCTCTCTGTTCAACAG 403
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 665 AGGCTTCTCTCTCTGAGAGCTCTCTCTGAGAGCTCTCTCTCTCTCTCTCTCTCT 702
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 CTCTCTCTCTCTGTTCAAGTAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 441
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: December 16, 2003, 17:31:45
Job time : 1452.89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 15:10:30 ; Search time 2514.24 Seconds

(without alignments)
3823.721 Million cell updates/sec

Title: US-09-864-291-12
Sequence: 1 MPPDLMNTLTYEGVPFANF.....AQAPENRASPSSSQVHS 235

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888771 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgm2.1/USPTO.spool/p/US09864291/runat.15122003.160858.23305/app.query.fasta.1.846
-DB=genmb1 -GEMT=fastap -SUPERfix=xge -MINMATCH=0.1 -LOOPTCL=0 -LOOPTXT=0
-UNITs=bits -STARF=1 -END=1 -MATRIX=biocum62 -TRANS=humand40 cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODS=LOCAL
-OUTFMT=ptc -NORM=ext -HSAFSLZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09864291.qcgn.1.1.3854.qrunat.15122003.160858.23305 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESUBSTR -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database : Genmb1:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vt:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rdt:*
36: em_hcg_man:*
37: em_hcg_vrt:*
38: em_ey:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1279	100.0	1001	6 AX359665	AX359665 Sequence
2	1263	98.7	2267	9 BC022546	BC022546 Homo sapi
3	1235	96.6	2266	9 BC022549	BC022549 Homo sapi
4	777	60.8	1413	6 AX359658	AX359658 Sequence
5	765	59.8	220895	9 HS250D10	Z59716 Human DNA s
6	429	33.5	168425	2 AC113593	AC113593 Mus muscu
7	429	33.5	224086	10 AC104325	AC104325 Mus muscu
8	424	33.2	179222	2 BX296515	BX296515 Sus scrofa
9	424	33.2	222469	2 BX470149	BX470149 Dario rer
10	405	31.7	251149	2 AC107527	AC107527 Rattus no
11	405	31.7	270171	2 AC132969	AC132969 Rattus no
12	383	29.9	129624	2 BX296540	BX296540 Sus scrofa
13	289	22.6	170781	2 AC092212	AC092212 Trypanoso
14	276.5	21.6	2170	9 AB072784	AB072784 Macaca fa
15	272.5	21.3	166045	2 AC134957	AC134957 Rattus no
16	270	21.1	793	10 AF459026	AF459026 Rattus no
17	261.5	20.4	1802	10 MM040826	U40826 Mus muscicu
18	261.5	20.4	1885	6 AR220846	AR220846 Sequence
19	258.5	20.2	198276	2 AC145040	AC145040 Gorilla g
20	258	20.2	196322	2 AC145025	AC145025 Gorilla g
21	257	20.1	1691	9 HS079458	U79458 Human WW do
22	257	20.1	1820	9 BC007452	BC007452 Homo sapi
23	257	20.1	1835	9 AK057881	AK057881 Homo sapi
24	257	20.1	1912	9 BC010616	BC010616 Homo sapi
25	256	20.0	148851	9 HS155D22	Z97205 Human DNA s
26	255.5	20.0	198833	2 AC144590	AC144590 Gorilla g
27	254.5	19.9	207682	2 AC144462	AC144462 Gorilla g
28	253	19.8	1747	3 AY060828	AY060828 Drosophi
29	252.5	19.7	2557	8 AF255446	AF255446 Cryptheco
30	251	19.6	174921	2 AC012266	AC012266 Homo sapi
31	250	19.5	167067	10 AC124489	AC124489 Mus muscu
32	249.5	19.5	39782	2 AC138635	AC138635 Homo sapi
33	249.5	19.5	40476	2 AC140538	AC140538 Homo sapi
34	244.5	19.1	159064	9 AC138512	AC138512 Homo sapi
35	244	19.1	192533	9 AC055876	AC055876 Homo sapi
36	242.5	19.0	2557	8 AF255446	AF255446 Cryptheco
37	240.5	18.8	198438	2 AC144362	AC144362 Gorilla g
38	239	18.7	151800	2 AC138649	AC138649 Homo sapi
39	236.5	18.5	160174	2 AC091951	AC091951 Homo sapi
40	236.5	18.5	182255	2 AC140810	AC140810 Homo sapi
41	235.5	18.4	199175	2 AC140836	AC140836 Homo sapi
42	234.5	18.3	1205	5 BC053232	BC053232 Dario rer
43	233.5	18.3	141447	9 AC016204	AC016204 Homo sapi
44	233	18.3	180233	9 AC011767	AC011767 Homo sapi
45	232	18.1	235381	2 AC103031	AC103031 Rattus no

ALIGNMENTS

RESULT 1

AX359665	LOCUS	AX359665	1001 bp	DNA	linear	PAT 13-FEB-2002
AX359665	DEFINITION	Sequence 11 from Patent WO0190185.				
AX359665	ACCESSION	AX359665				
AX359665.1	VERSION	GI:18675411				
AX359665	KEYWORDS					
SOURCE	ORGANISM	Homo sapiens (human)				
REFERENCE	AUTHORS	Okamoto, R. and Sutovsky, P.				
TITLE		Pt23 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses thereof				
JOURNAL		Patent: WO 0190185-A 11 29-NOV-2001;				
		QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES UNIVERSITY (US)				
FEATURES	source	Location/Qualifiers				
		1..1001				
		/organism="Homo sapiens"				
		/mol_type="Genomic DNA"				
		/db_xref="taxon:9606"				
CDS		1..708				
		/note="unannotated protein product"				
		/codon_start=1				
		/protein_id="CAD23337.1"				
		/db_xref="GI:18675412"				
		/translation="MPPDLMTNLTVBOVPANAFIKGTIOAAPGYGMEGATPLVFP				
		NGDAIRFQALVTKASAVARGPRLTMDPMSNGIYVIRGNMCTPDMCSVYVGG				
		APPAQIGAPPPGYGAPPPAGYGAQPPVGNPVPYRAGAPPLGAPPAQYGAQ				
		PLGGAAPPLGYGTGPPPLGYGAPPLGYGAPPAQNGPPAGPAGPAGPAGPAGP				
		PENRSLPSASSSQVHS"				
BASE COUNT	261 a	254 c	248 g	238 t		
ORIGIN						
Alignment Scores:						
Pred. No.:	1.75e-50	Length:	1001			
Score:	1279.00	Matches:	235			
Percent Similarity:	100.00%	Conservative:	0			
Best Local Similarity:	100.00%	Mismatches:	0			
Query Match:	100.00%	Indels:	0			
DB:	6	Gaps:	0			
US-09-864-291-12 (1-235) x AX359665 (1-1001)						
QY	1	MetProPheAepLeuMetThzAsnLeuThrValGluGlnProValPheAlaAlaAsn	20			
DB	1	ATGCCATTGATCTGATGACGAACTCTACTGTGGAACAACAGTATTTCCTGCAAACTTC	60			
QY	21	IlleuysgIyThrlleGlnAlaAlaProTyrgIyITPglugIyGlnAlaThrPhelys	40			
DB	61	ATTAGGGAACATATTCAAGCAGCTCCATATGCTGGCTGGGAAGCAAGCTTCTTTAA	120			
QY	41	LeuValPheAysAysGlyAAspAlalleGluPheAlaGlnIleuMetVallyAlaAlaSer	60			
DB	121	TTAGCTTCAGAAATGAGAGATCCATGATTTGCCAGTTGATGTGAAGCTCTCT	180			
QY	61	AlaValAlaAysGlyPheProLeuAysThzrLeuAysAepTTPPheSerSerMetGlyIle	80			
DB	181	GCTGTGGCCCGAGATTTCCACTTGAACCTTAATATGACTGGTTTCAGCTCATGGGAAT	240			
QY	81	TyrValIleIethrIyGluGlyAAsnMetCysThrProGlnMetProCysSerValIleVal	100			
DB	241	TATGTAATTAATCGGGGAAGGGAATATGTCACCTCCAGATGCTTGTTCAGTTATTTGTC	300			
QY	101	TyrgIyAlaProProAlaGlyTyrgIyAlaProProGlyTyrgIyAlaProProAla	120			
DB	301	TATGGGGCCCACTCGCAGATATATGAGCCCACTCCCGATACGGAAGCCCACTGCA	360			
QY	121	GlyTyrgIyAlaGlnProValGlyAAsnGluGlyProProValGlyTyrrAysAlaSerPro	140			
DB	361	GGATTTGAGCCCAACCCGTAGGAATGAAAGGCCCGCTGTGGGATTAAGAGCTCACT	420			

QY		141	ValArglyrTgIyAlaProFroLeugLyTgIyAlaProFroLaagiYTgIyAlaPro	160
Db		421	GtGGAGATATGAGGCCCACTCTTGATATGAGGCCCACTCGAGATATGAGCCCA	480
QY		161	ProLeugLyTgIyAlaProFroLeugLyTgIyThrProFroLeugLyTgIyAla	180
Db		481	CCTTAGAATATGAGGCCCACTCTTGATATGAAACCCCACTTCGATATGAGACC	540
QY		181	ProFroLeugLyTgIyAlaProFroLaagiYAasnIunLyProFroLaagiYTgAr	200
Db		541	CCACTTCGAATATGAGGCCCACTCGAGAATATAAGCCCGCTCGGGATTACA	600
QY		201	AlaseerProLaagiSerGIyAlaArProGIInuSerThrAlaaGIAnaLaproGIu	220
Db		601	GCCCTCACTCTGATATGAGGCCCACTCGAATATCAGACAGCCCAAGCTCTGA	660
QY		221	AsnGIuAlaSerIeuProSerAlaSerSerSeriAlaHiser	235
Db		661	AACGAGGCTTCTCTCCCTCGCTCTCTTCTCAGGTCAATCT	705
RESULT 2				
BC022546				
LOCUS				
DEFINITION			Homo sapiens, similar to RIKEN cDNA 4930521I23 gene, clone	
ACCESSION			MGC:26816 IMAGE:4811804, mRNA, complete cds.	
VERSION			BC022546.1 GI:18490706	
KEYWORDS			MGC.	
SOURCE			Homo sapiens (human)	
ORGANISM			Homo sapiens	
REFERENCE			Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLS			1 (bases 1 to 2267)	
JOURNAL			Strausberg,R. Direct Submision Submitted (01-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	
REMARK			NIH-MGC Project URL: http://mgc.nci.nih.gov	
COMMENT			Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shitaki Toshiyuki and Piero Carninci (RIKEN) cDNA library Arrayed by: The I.M.A.G.E. Consortium (ILNLT) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@packil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.	
FEATURES				
source				
			location/Qualifiers	
			1..2267	
			/organism="Homo sapiens"	
			/mol_type="mRNA"	
			/db_xref="taxon:9606"	
			/clone="MGC:26816 IMAGE:4811804"	
			/tissue type="Brain, hippocampus"	
			/clone_id="NIH_MGC_95"	
			/lab_host="DHIOB"	
			/note="Vector: pBluescript"	
			32..961	
			/codon_start=1	
			/product="Similar to RIKEN cDNA 4930521I23 gene"	
			/protein_id="AAH22546.1"	
CDS				

```

/db_xref="GI:18490710"
/translation="MAVNSHTENRRGALIPNGESILKRSPNVELSPPOQSEGSNVFS
GRKGTGLFTLSYRVIFITSCSIDPMISFMMPLMTNLTVBOVPFANPKGTIOA
PYGMEGQATKLVFRNGAIBFQALVKASAAARGPPLATLNDWFSMSKIYITIG
GNMCTPQMPCSVIVGAPPAVGPAGVAPPGVAGQAPVAGNPGPPVGVAPVGTG
APPLGAGAPPAVGPAGVAPPGVAGQAPVAGNPGPPVGVAPVGTGAPPAVGPAGV
PAGGAPRHSSTAAOPENEAALPSASSQVHS"
BASE COUNT      709 a      477 c      469 g      612 t
ORIGIN
Alignment Scores:
Pred. No.:      1,81e-49      Length:      2267
Score:          1263.00      Matches:      232
Percent Similarity: 98.72%      Conservative: 0
Best Local Similarity: 98.72%      Mismatches: 3
Query Match:    98.75%      Indels:      0
Gaps:           9           Gaps:      0
US-09-864-291-12 (1-235) x BC022546 (1-2267)
QY      1 MetProheapleumMetThrasuleuthThValGlugInProvalPhealAlaAanPhe 20
D      254 ATGCCATTGATTCGATGACGAACTCACTGTGAACAACAGATTTGCTGCAAACTTC 313
QY      21 IlleYsgIYThriIlegInAlaAProTyrgIYgIYTPgIugIYgInAlaThPhelYs 40
D      314 ATTAAGGAACTATTTCAGGACGCTCATATGCTGAGGAGACAGACTTCTTTAA 373
QY      41 LeuValPheArGsnGlyAspAlaIlegIuphealagInleuMetValAlaAaser 60
D      374 TTAGCTTCAGAAATGAGAGGCGCTTGAATTTGCCCGATTGATGATTAAGCTCCCT 433
QY      61 AlaValAlaArGlyPheProleuArGThrleuAanArPTTPheSerSerMetGlyLe 80
D      434 GCTGTCGCCAGAGATTCCACTTGAACCTTAATGACTGCTTATGGAAT 493
QY      81 TyrValIleThnGlyIugIYAsnMetCyThrProGInMetProCySerValIleVal 100
D      494 TATGTAATTAATCTGGGAGGAAATGTCACCTCAAGATGCTTGTTCAGTTATTTTC 553
QY      101 TyrGlyAlaProProAlaGlyTYrgIYAlaProProGlyTYrgIYAlaProProAla 120
D      554 TATGAGGCCCACTCGAGATATGAGCCCACTCCCGATACGAGGCCCACTTCA 613
QY      121 GlyTYrgIYAlaInProValGlyAanGlyIYProProValGlyTYrAlaAaserPro 140
D      614 GGAATATGAGGCCCAACCGTGAATATGAGCCCGCTGAGATACAGGCTCACT 673
QY      141 ValArGlyTYrgIYAlaProProleuGlyTYrgIYAlaProProAlaGlyTYrgIYAlaPro 160
D      674 GTGCGATATGAGGCCCACTCTTGATACGAGGCCCACTCGAGATATGAGGCCCA 733
QY      161 ProleuGlyTYrgIYAlaProProleuGlyTYrgIYThrProProleuGlyTYrgIYAla 180
D      734 CCTTGAAGATATGAGGCCCACTCTTGATATGAGGCCCACTCTCGAGATATGAGGCC 793
QY      181 ProProleuGlyTYrgIYAlaProProAlaGlyAanGlyIYProProAlaGlyTYrAla 200
D      794 CCACTCTCGAGATATGAGGCCCACTCGAGATATGAGGCCCGCTCGCGAGATACGA 853
QY      201 AlaserProAlaGlySerGlyAlaArGProGInGlySerThrAlaIglInAlaProGlu 220
D      854 GCCTCACTCTGCTGATCAAGAGCCAGGCTTCAAGATCAAGAGCCAGGCTCTTAA 913
QY      221 AenGluAlaserleuProserAlaserSerSerGluValIaser 235
D      914 AACGAGGCTCTCTCTCCCTGCTCTCTCTCTCAAGTCAATTC 958
RESULT 3
BC022549      2266 bp      mRNA      linear      PRI 04-FEB-2002
LOCUS      Homo sapiens, similar to RIKEN cDNA 4930521I23 gene, clone
DEFINITION      MGC:26828 IMAGE:4815849, mRNA, complete cds.

```

```

ACCESSION      BC022549
VERSION        BC022549.1
KEYWORDS       MGC.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
REFERENCE      1 (bases 1 to 2266)
AUTHORS        Strausberg, R.
TITLE          Direct Submission
JOURNAL        Submitted (01-FEB-2002) National Institutes of Health, Mammalian
               Gene Collection (MGC), Cancer Genomics Office, National Cancer
               Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
               USA
REMARK         NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT         Contact: MGC help desk
               Email: gcgaps-remail.nih.gov
               Tissue Procurement: Miklos Palokvite, M.D., Ph.D.
               cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
               Tohyuki and Piero Carninci (RIKEN)
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
               DNA Sequencing by: Sequencing Group at the Stanford Human Genome
               Center, Stanford University School of Medicine, Stanford, CA 94305
               Web site: http://www-shgc.stanford.edu
               Contact: (Dickson, Mark) mcd@paxil.stanford.edu
               Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
               R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAC Plate: 32 Row: m Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: similarity but not
identity to protein.
Location/Qualifiers
1..2266
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26828 IMAGE:4815849"
/tissue_type="Brain, hippocampus"
/clone_id="NIH MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
32..634
/codon_start=1
/product="Similar to RIKEN cDNA 4930521I23 gene"
/protein_id="AAH2549.1"
/db_xref="GI:18490710"
/translation="MAVNSHTENRRGALIPNGESILKRSPNVELSPPOQSEGSNVFS
GRKGTGLFTLSYRVIFITSCSIDPMISFMMPLMTNLTVBOVPFANPKGTIOA
PYGMEGQATKLVFRNGAIBFQALVKASAAARGPPLATLNDWFSMSKIYITIG
GICAHKCLIVQLLSMBPHLDMPHPLDTEPHLDMPNP"
BASE COUNT      709 a      477 c      469 g      612 t
ORIGIN
Alignment Scores:
Pred. No.:      3.38e-48      Length:      2266
Score:          1235.00      Matches:      231
Percent Similarity: 98.30%      Conservative: 0
Best Local Similarity: 98.30%      Mismatches: 4
Query Match:    96.56%      Indels:      1
Gaps:           9           Gaps:      0
US-09-864-291-12 (1-235) x BC022546 (1-2266)
QY      1 MetProheapleumMetThrasuleuthThValGlugInProvalPhealAlaAanPhe 20
D      254 ATGCCATTGATTCGATGACGAACTCACTGTGAACAACAGATTTGCTGCAAACTTC 313
QY      21 IlleYsgIYThriIlegInAlaAProTyrgIYgIYTPgIugIYgInAlaThPhelYs 40
D      314 ATTAAGGAACTATTTCAGGACGCTCATATGCTGAGGAGACAGACTTCTTTAA 373

```


gene similar to neuronal-specific septin 3, a pseudogene similar to AN2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a genomic marker D22S1178, a CA repeat polymorphism, ESTs and a CpG island, complete sequence.

ACCESSION 299716 GI:4456457
 VERSION 299716.4
 KEYWORDS HTG; AN2; CpG Island; D22S1178; NAGA; septin 3; SRBP2.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 220895)

REFERENCE 1 (bases 1 to 220895)
 AUTHORS Clark, G.
 TITLE Direct Submission
 JOURNAL Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

COMMENT On Mar 21, 1999 this sequence version replaced gi:4164339. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMBSP; Information on the WORMBSP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormbsp CTA-250D10 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.

VECTOR: pBAC108L

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed from the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>
 This sequence is the entire insert of clone CTA-250D10. The true left end of clone RP1-18601 is at 129979 in this sequence. The true right end of clone RP5-821D11 is at 23458 in this sequence. The true right end of clone RP3-359J16 is at 118711 in this sequence.

FEATURES

SOURCE

1. 220895
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="CTA-250D10"
 /clone_1lb="CIT978SK-A1"
 17. 29
 /note="2.6 copies 5 mer ATTGG 26% conserved"
 repeat_region
 33. 3325
 /note="AluX repeat: matches 1. 301 of consensus"
 repeat_region
 326. 338
 /note="2.6 copies 5 mer TTGGA 26% conserved"
 repeat_region
 667. 683
 /note="2.1 copies 8 mer AGAGACC 34% conserved"
 repeat_region
 740. 751
 /note="2.4 copies 5 mer TTTTA 24% conserved"

repeat_region 1443. 1457
 /note="7.5 copies 2 mer AC 21% conserved"
 repeat_region 1715. 1997
 /note="AluO repeat: matches 5. 293 of consensus"
 repeat_region 2103. 2378
 /note="AluSg repeat: matches 2. 296 of consensus"
 repeat_region 2844. 2855
 /note="12.0 copies 1 mer A 24% conserved"
 repeat_region complement(2884. 3033)
 /note="MIR repeat: matches 70. 235 of consensus"
 repeat_region complement(3040. 3344)
 /note="AluY repeat: matches 1. 310 of consensus"
 repeat_region 3446. 3681
 /note="L1MB4a repeat: matches 5256. 5471 of consensus"
 repeat_region 3746. 3837
 /note="L1MB4a repeat: matches 5615. 5707 of consensus"
 repeat_region 3838. 4142
 /note="AluO repeat: matches 2. 305 of consensus"
 repeat_region 4143. 4441
 /note="L1MB4a repeat: matches 5707. 5939 of consensus"
 repeat_region 4442. 4747
 /note="AluSg repeat: matches 1. 292 of consensus"
 repeat_region 4748. 4858
 /note="L1MB4a repeat: matches 5999. 6110 of consensus"
 repeat_region 4881. 5172
 /note="AluX repeat: matches 1. 296 of consensus"
 repeat_region 5418. 5429
 /note="2.4 copies 5 mer GCTGA 24% conserved"
 repeat_region complement(5769. 6021)
 /note="match: G8S: Em:A0069839"
 repeat_region 6074. 6089
 /note="2.0 copies 8 mer CCGGCTG 32% conserved"
 repeat_region 6139. 6153
 /note="2.5 copies 6 mer TTTAA 21% conserved"
 repeat_region 6141. 6155
 /note="3.0 copies 5 mer TTTAA 21% conserved"
 repeat_region complement(6366. 6572)
 /note="MIR repeat: matches 5. 214 of consensus"
 repeat_region 6839. 6852
 /note="2.3 copies 6 mer GGAAT 28% conserved"
 repeat_region 7088. 7114
 /note="6.8 copies 4 mer TTTA 36% conserved"
 repeat_region 7091. 7109
 /note="1.9 copies 10 mer ATTATTTAC 38% conserved"
 repeat_region complement(7122. 7418)
 /note="AluX repeat: matches 1. 300 of consensus"
 repeat_region 7464. 7502
 /note="19.5 copies 2 mer AT 78% conserved"
 repeat_region 7502. 7524
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region 7523. 7538
 /note="2.0 copies 8 mer CAAATAT 32% conserved"
 repeat_region 7535. 7563
 /note="14.5 copies 2 mer AT 49% conserved"
 repeat_region complement(7564. 7859)
 /note="AluY repeat: matches 1. 296 of consensus"
 repeat_region 7873. 7893
 /note="5.2 copies 4 mer TGT 33% conserved"
 repeat_region 7914. 8200
 /note="AluO repeat: matches 1. 290 of consensus"
 repeat_region complement(8202. 8318)
 /note="AluO/FLM repeat: matches 1. 117 of consensus"
 repeat_region 8387. 8409
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region 8483. 8498
 /note="2.0 copies 8 mer GTGTGTGT 32% conserved"
 repeat_region 8486. 8496
 /note="2.2 copies 5 mer GTGCT 22% conserved"
 repeat_region 8606. 8622
 /note="2.4 copies 7 mer AGAGAA 34% conserved"
 repeat_region complement(8767. 8989)
 /note="MER115 repeat: matches 33. 268 of consensus"
 repeat_region 9004. 9014


```

repeat_region
/note="2.2 copies 5 mer CCTGG 224 conserved"
9037..9140
/note="MIR repeat: matches 17..123 of consensus"
repeat_region
9509..9823
/note="AluX repeat: matches 1..312 of consensus"
9872..9883
/note="2.4 copies 5 mer GTCTT 244 conserved"
gene
10071..49606
/gene="SREBF2"
join(<10071..10520,11851..12032,14129..14275,17038..17259,
18577..18691,18783..18964,20469..20661,21182..21363,
23956..24232,28082..28251,36357..36525,38060..38177,
40292..40401,41889..42021,43570..43738,46210..46395,
48103..48214,48680..49606)
/gene="SREBF2"
/product="bK250D10.1 (sterol regulatory element binding
transcription factor 2)"
/note="sterol regulatory element binding transcription
factor 2"
match: CDNA: Bm:U02031 Bm:U12330 Bm:U00968 Bm:U09103
Bm:U16995 Bm:U22818 Bm:U22819 Bm:U12329 Bm:AB017337
match: ESTs: Bm:R36312 Bm:AL040138 Bm:H92808 Bm:AA05886"
CDS
/evidence="not_experimental"
join(<10071..10520,11851..12032,14129..14275,17038..17259,
18577..18691,18783..18964,20469..20661,21182..21363,
23956..24232,28082..28251,36357..36525,38060..38177,
40292..40401,41889..42021,43570..43738,46210..46395,
48103..48214,48680..48900)
/gene="SREBF2"
/note="match: proteins: Sw:Q12772 Sw:Q60429 Sw:P36956
Sw:Q60416 Tr:O15352"
/codon_start=3
/evidence="not_experimental"
/product="bK250D10.1 (sterol regulatory element binding
transcription factor 2)"
/protein_id="CA841234.1"
/db_xref="GI:5578715"
/translation="MLQFVSNVGVGPPDLFSEDLCSPPSGSGSSSGSSSSSS
NKGSSGAVDSVPSVPTQVTLSPSPSAAPALPTLONATSPSPPTPAPLILQ
PRQPPQPPQTOLOQOTWITTFSTTPTRITIQPLITONATSPSPQVLOVQOBIYT
SSQVQPVITQQQVTOQAVLQOTANGLQTLAPATVGTAAVQVQVAPVQVQVPII
KDSVLATLTKDGSFVMAVONPALTALPTQALASVPTLVGSGGILITPMAM
GGRKVIKQVGGVQVLEPPKGGERTTHNIEKRRSSINDIKILQDLMVTDAM
HKSGLVAKAIDYIKYLOQVNHKLROEMVTKLANOKMLKGLIDSLVDNEKDLIS
DRQNVTLMSPPASDSQAGSPRSISIDSPGSLDADKAVDPSPSPVVALCMVRS
KILCVLPLCLSFNPLTSLQMGASHDSQHPHSGSVLSFESGSGFPMWMT
LILMIVGVIVLSVFKYLVHGEPVIRPHRSSVTWRHKKQDLIDAGDFPAAAGN
LQCLAVIGRALPTSLDLACSLSMVIRYSLOKRLVRLVTLKVDQCRATPAAAG
FDEATSTADALAVHRLHQLHITGKLPAQSACSDVHMAICAVNLACAEKIPST
LVRIHLTAMGLTKRCGKLGFLASFLSRAQSLCGPESNAVPDSLRMLCHPLGQKFP

```

Alignment Scores:

Pred. No.:	Length:	220895
Score:	765.00	Matches: 138
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	59.81%	Indels: 0
DB:	9	Gaps: 0

US-09-864-291-12 (1-235) x HS250D10 (1-220895)

QY 98 ValIleValTyrGlyAlaProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAla 117

DB 170005 GTTATGTCATGAGAGCCCACTGACGATATGAGACCCCACTCCGATACGAGGCC 170064

QY 118 ProProAlaGlyTyrGlyAlaGlnProValaGlyAsnGluGlyProProAlaGlyTyrArg 137

DB 170065 CCACTGACGATATGAGAGCCCACTGAGAAATGAAAGCCCGCTGAGGATACAGA 170124

QY 138 AlSerProValaGlyTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyr 157

DB 170125 GCCTCACTGTGAGATGAGAGCCCACTCTTGATACGAGAGCCCACTGACGATAT 170184

QY 158 GlyAlaProProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGly 177

DB 170185 GGAGCCCACTCTGATATGAGAGCCCACTCTTGATATGAGAGCCCACTCTCGGA 170244

QY 178 TyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAla 197

DB 170245 TATGAGAGCCCACTCTCGAATGAGAGCCCACTGACGAAATGAAAGCCCGCTGG 170304

QY 198 GlyTyrArgAlaSerProAlaGlySerGlyAlaArgProGlnGlySerThrAlaAlaGln 217

DB 170305 GATATACAGAGCTCACTGCTGATCAGAGAGAGAGCTCAAGAAATCTACAGAGCCGAG 170364

QY 218 AlaProGluAsnGluAlaSerLeuProSerAlaSerSerGlnValHisSer 235

DB 170365 GCTCCGAAAAGAGGCTCTCTCCCTCTGCTCTCTCTCTCTCTCTCTCTCTCT 170418

RESULT 6

AC113593

LOCUS

DEFINITION

Mus musculus clone RP23-363124, WORKING DRAFT SEQUENCE, 18 ordered pieces.

AC113593

AC113593.3 GI:21327432

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 168425)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Mus musculus, clone RP23-363124

Unpublished

2 (bases 1 to 168425)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckgatter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepl, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Lander, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meidrin, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Nguyen, C., Nicol, R., Nordu, C., Norman, C.H., O'Connor, T., O'Neill, D., Oliver, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., O'Neill, J., Peterson, K., Phunhang, P., Pierre, N., Pollara, V., Raymond, C., Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnback, R., Seaman, S., Severy, P., Spencer, B., Stange, Thomann, N., Stojanovic, N., Strahan, N., Sudramanlan, A., Talamas, J., Testaye, S., Theodore, J., Topnam, K., Travers, M., Traves, N., Triggillo, J., Vassiliev, J., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (03-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 168425)

Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckgatter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepl, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzderald, M., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Lander, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meidrin, J., Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Nguyen, C., Nicol, R., Nordu, C., Norman, C.H., O'Connor, T., O'Neill, P.,

TITLE JOURNAL COMMENT

O'Neill, D., Oliver, J., Peterson, K., Phunhphang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Riese, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Sudratanian, A., Talama, J., Testaye, S., Theodore, J., Topham, K., Travers, N., Travis, N., Triggillo, J., Vassiliev, H., Vei, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J., Young, G., Zainoun, J., Zemdek, L., Zimer, A. and Zody, M.

Submitted (06-JUN-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced g1:21313869.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/BM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: I23816
Center clone name: 363_1_24

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 162227 bases at least Q40
Consensus quality: 164865 bases at least Q30
Consensus quality: 165972 bases at least Q20
Insert size: 162000; agarose-fp
Insert size: 166725; sum-of-coverage
Quality coverage: 8.6 in Q20 bases; sum-of-coverage
Quality coverage: 8.3 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1 1412: contig of 1412 bp in length
1413 1512: gap of 100 bp
1513 3366: contig of 1854 bp in length
3367 3466: gap of 100 bp
3467 5542: contig of 2076 bp in length
5543 5642: gap of 100 bp
5643 7949: contig of 2307 bp in length
7950 8049: gap of 100 bp
8050 42610: contig of 34561 bp in length
42611 42710: gap of 100 bp
42711 45717: contig of 3007 bp in length
45718 45817: gap of 100 bp
45818 50106: contig of 4289 bp in length
50107 50206: gap of 100 bp
50207 52945: contig of 2739 bp in length
52946 53045: gap of 100 bp
53046 56263: contig of 3224 bp in length
56264 56369: gap of 100 bp
56370 56987: contig of 3528 bp in length
56988 59997: gap of 100 bp
59998 67207: contig of 7210 bp in length
67208 67307: gap of 100 bp
67308 78743: contig of 11436 bp in length
78744 78843: gap of 100 bp
78844 89991: contig of 11148 bp in length
89992 90091: gap of 100 bp
90092 107918: contig of 17827 bp in length
107919 108018: gap of 100 bp
108019 124874: contig of 16856 bp in length

124875 124974: gap of 100 bp
124975 142093: contig of 17119 bp in length
142094 142193: gap of 100 bp
142194 167111: contig of 24918 bp in length
167112 167212: gap of 100 bp
167212 168425: contig of 1214 bp in length.
Location/Qualifiers
1. 168425
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-363124"
/clone_1lb="RP23-363124"
1. 1412
/note="assembly_fragment"
1513. 3366
/note="assembly_fragment"
3467. 5542
/note="assembly_fragment"
5643. 7949
/note="assembly_fragment"
8050. 42610
/note="assembly_fragment"
42711. 45717
/note="assembly_fragment"
45818. 50106
/note="assembly_fragment"
50207. 52945
/note="assembly_fragment"
53046. 56269
/note="assembly_fragment"
56370. 56987
/note="assembly_fragment"
59998. 67207
/note="assembly_fragment"
67308. 78743
/note="assembly_fragment"
78844. 89991
/note="assembly_fragment"
90092. 107918
/note="assembly_fragment"
108019. 124874
/note="assembly_fragment"
124975. 142093
/note="assembly_fragment"
142194. 167111
/note="assembly_fragment"
167212. 168425
/note="assembly_fragment"
vector_end:17
vector_side:right

BASE COUNT 44391 a 39572 c 38747 g 44015 t 1700 others
ORIGIN

Alignment Scores:

Pred. No.: 4,41e-10 Length: 168425
Score: 429.00 Matches: 94
Percent Similarity: 55.03% Conservative: 10
Best Local Similarity: 49.74% Mismatches: 37
Query Match: 33.54% Gaps: 48
DB: 2

US-09-864-291-12 (1-235) x AC113593 (1-168425)

QY 94 MetProCyseRerValIleValTYrGYAlaProProAlaGYrTYrGYAlaProPro 113

Db 66132 TRGCCATTTTCATT---GCCATGAGACCCCACTCGGAGATAGGGCTCACCCTG 66188

QY 114 GLYTYrGYAlaProProAlaGYrTYrGYAlaGlnProValGYrYanGluGYrProPro 133

Db 66189 GAGTACGAGATCCCATCTGACAGGATAGGAGCCCACTCGGAGTACGAGCCGCCA 66248

QY 134 ValGYrTYrGYAla---SerPro-----Val 141

COMMENT On May 7, 2003 this sequence version replaced gl:30387077.

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information
 Center project name: ZK242K7
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 197958 bases at least Q40
 Consensus quality: 206715 bases at least Q30
 Consensus quality: 212351 bases at least Q20
 Insert size: 217569; sum-of-contigs
 Insert size: 165837; 6.2% error; agarose-fp
 Quality coverage: 2.48x in Q20 bases; sum-of-contigs Quality
 coverage: 3.95x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 50 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      6656: contig of 6656 bp in length
*      6657      6756: gap of 100 bp
*      6757      10255: contig of 3499 bp in length
*      10256      10355: gap of 100 bp
*      10356      13223: contig of 2868 bp in length
*      13224      13323: gap of 100 bp
*      13324      19935: contig of 6612 bp in length
*      19936      20035: gap of 100 bp
*      20036      22597: contig of 2562 bp in length
*      22598      22697: gap of 100 bp
*      22698      28569: contig of 5872 bp in length
*      28570      28669: gap of 100 bp
*      37965      37966: contig of 9296 bp in length
*      37966      38065: gap of 100 bp
*      38066      41140: contig of 3075 bp in length
*      41141      41240: gap of 100 bp
*      41241      45108: contig of 3868 bp in length
*      45109      45208: gap of 100 bp
*      45209      47730: contig of 2522 bp in length
*      47731      47830: gap of 100 bp
*      47831      51884: contig of 4054 bp in length
*      51885      51984: gap of 100 bp
*      51985      55857: contig of 3873 bp in length
*      55858      55957: gap of 100 bp
*      55958      60498: contig of 4541 bp in length
*      60499      60598: gap of 100 bp
*      60599      63641: contig of 3043 bp in length
*      63642      63741: gap of 100 bp
*      63742      72674: contig of 8933 bp in length
*      72675      72774: gap of 100 bp
*      72775      75602: contig of 2828 bp in length
*      75603      75702: gap of 100 bp
*      75703      78062: contig of 2260 bp in length
*      78063      78162: gap of 100 bp
*      78163      84485: contig of 6323 bp in length
*      84486      84585: gap of 100 bp
*      84586      87553: contig of 2968 bp in length
*      87554      87653: gap of 100 bp
*      87654      92487: contig of 4834 bp in length
*      92488      92587: gap of 100 bp
*      92588      96104: contig of 3517 bp in length
*      96105      96204: gap of 100 bp
*      96205      101479: contig of 5275 bp in length
*      101480      101579: gap of 100 bp
*      101580      104979: contig of 3400 bp in length
*      104980      105079: gap of 100 bp

```

```

*      105080      108003: contig of 2924 bp in length
*      108004      108103: gap of 100 bp
*      108104      113151: contig of 5048 bp in length
*      113152      113251: gap of 100 bp
*      113252      116585: contig of 3334 bp in length
*      116586      116685: gap of 100 bp
*      116686      119346: contig of 2661 bp in length
*      119347      119446: gap of 100 bp
*      119447      123125: contig of 3679 bp in length
*      123126      123225: gap of 100 bp
*      123226      131466: contig of 8041 bp in length
*      131467      131666: gap of 100 bp
*      131667      136193: contig of 4827 bp in length
*      136194      136293: gap of 100 bp
*      136294      139482: contig of 3189 bp in length
*      139483      139582: gap of 100 bp
*      139583      141875: contig of 2233 bp in length
*      141876      141975: gap of 100 bp
*      141976      144122: contig of 2147 bp in length
*      144123      144222: gap of 100 bp
*      144223      146968: contig of 2746 bp in length
*      146969      147068: gap of 100 bp
*      147069      149399: contig of 2331 bp in length
*      149400      149499: gap of 100 bp
*      149500      153963: contig of 4464 bp in length
*      153964      154063: gap of 100 bp
*      154064      156579: contig of 2516 bp in length
*      156580      156679: gap of 100 bp
*      156680      158928: contig of 2249 bp in length
*      158929      159028: gap of 100 bp
*      159029      164542: contig of 5514 bp in length
*      164543      164642: gap of 100 bp
*      164643      172054: contig of 7412 bp in length
*      172055      172154: gap of 100 bp
*      172155      175210: contig of 3056 bp in length
*      175211      175310: gap of 100 bp
*      175311      179379: contig of 4069 bp in length
*      179380      179479: gap of 100 bp
*      179480      182698: contig of 3219 bp in length
*      182699      182798: gap of 100 bp
*      182799      186779: contig of 3981 bp in length
*      186780      195626: contig of 8747 bp in length
*      195627      195726: gap of 100 bp
*      195727      199386: contig of 3670 bp in length
*      199387      199496: gap of 100 bp
*      199497      206412: contig of 6916 bp in length
*      206413      209619: contig of 3107 bp in length
*      209620      209719: gap of 100 bp
*      209720      219083: contig of 9364 bp in length
*      219084      219183: gap of 100 bp
*      219184      222469: contig of 3286 bp in length.

```

FEATURES

```

source
1..222469
  /organism="Danio rerio"
  /mol_type="genomic DNA"
  /db_xref="taxon:7955"
  /clone_id="DKFZ-242K7"
  /clone_lib="DantolKey"
  1..6656
    /note="assembly fragment:01359"
    fragment_chain:1
  6757..10255
    /note="assembly fragment:01385"
    fragment_chain:1
  10356..13223
    /note="assembly fragment:00592"
    fragment_chain:1
  13324..19935
    /note="assembly fragment:00237"
    fragment_chain:1
  20036..22597

```


QY 141 -----ValArgTyrGlyAlaPro-----ProLeu 148
DB 15227 TCTGATACGAGCTGCGAGATATGATCCCTCTCTATACGATAGTACCCCTATG 15168
QY 149 G|Y|T|G|Y|A|A|P|P|P|P|A|G|Y|T|Y|G|Y|A|A|P|P|P|P|A|P|P|P|P| 168
DB 15167 GGCCTATGAGTCCCACTCTCTGATATGACCCCACTCTGATATGACGATCCCACT 15168
QY 169 LeuGlyTyr----- 171
DB 15107 CCTGATACGAGCCCACTATGAGTATGGGCCCCCACTCTATACGATAGTACCA 15048
QY 172 -----G|Y|T|H|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P| 187
DB 15047 CCTATGATCCGCGTCCCACTCTCTGATATGAGCCCACTATGAGTATGAGTATG 14988
QY 188 P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P| 200
DB 14987 CCTGCTCCTGAGGATGATGATCCTCTCTGATATGATCCTCTCTGATATGATG 14928
QY 201 A|A|S|E|P|P|A|G|Y|S|E|R|G|Y|A|A|A|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P| 220
DB 14927 GCACATACCTGCTGCTGCTGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 14868
QY 221 Aaag|A|A|S|E|L|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P|P| 235
DB 14867 TTTCAAGGCACTTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14823
RESULT 12
BX296540/c 129624 bp DNA 1linear HNG 02-APR-2003
LOCUS Sus scrofa clone P18r-231K18, *** SEQUENCING IN PROGRESS ***, 37
DEFINITION unordered pieces.
ACCESSION BX296540
VERSION BX296540.2 GI:29500961
KEYWORDS HNG, HNGS_PHASE1.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 129624)
AUTHORS Burton, J.
TITLES Direct Submision
JOURNAL Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Apr 2, 2003 this sequence version replaced gi:29335441.
COMMENT ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BT231K18
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Chemistry: Dye-terminator, 100% of reads
Consensus quality: 113740 bases at least Q40
Consensus quality: 118187 bases at least Q30
Consensus quality: 120712 bases at least Q20
Insert size: 126024, sum-of-coverage
Insert size: 154528, 4.6% error; agarose-fp
Quality coverage: 2.35x in Q20 bases; sum-of-coverage Quality
coverage: 3.19x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 4262: contig of 4262 bp in length
* 4263 gap of 100 bp
* 4363 contig of 2629 bp in length
* 6992 gap of 100 bp
* 7092 contig of 2389 bp in length
* 9481 gap of 100 bp
* 9581 contig of 2108 bp in length
* 11689 gap of 100 bp
* 11789 contig of 2870 bp in length
* 14659 gap of 100 bp
* 14759 contig of 3682 bp in length
* 18441 gap of 100 bp
* 18541 contig of 2278 bp in length
* 20819 gap of 100 bp
* 20919 contig of 2051 bp in length
* 22970 gap of 100 bp
* 23070 contig of 4515 bp in length
* 27585 gap of 100 bp
* 27685 contig of 5527 bp in length
* 33212 gap of 100 bp
* 33312 contig of 2210 bp in length
* 35522 gap of 100 bp
* 35622 contig of 2207 bp in length
* 38529 gap of 100 bp
* 38629 contig of 3435 bp in length
* 42064 gap of 100 bp
* 42164 contig of 2001 bp in length
* 44165 gap of 100 bp
* 44265 contig of 3443 bp in length
* 47708 gap of 100 bp
* 47808 contig of 2930 bp in length
* 50737 gap of 100 bp
* 50837 contig of 6229 bp in length
* 50838 gap of 100 bp
* 57067 contig of 2812 bp in length
* 57167 gap of 100 bp
* 59978 contig of 2482 bp in length
* 60079 gap of 100 bp
* 62561 contig of 2203 bp in length
* 62661 gap of 100 bp
* 64964 contig of 2128 bp in length
* 65064 gap of 100 bp
* 67192 contig of 4714 bp in length
* 67292 gap of 100 bp
* 72006 contig of 4490 bp in length
* 72105 gap of 100 bp
* 76595 contig of 3130 bp in length
* 76995 gap of 100 bp
* 76996 contig of 3615 bp in length
* 79825 gap of 100 bp
* 83540 contig of 2636 bp in length
* 83541 gap of 100 bp
* 86276 contig of 8244 bp in length
* 86376 gap of 100 bp
* 86377 contig of 8244 bp in length
* 94620 gap of 100 bp
* 94720 contig of 2800 bp in length
* 94721 gap of 100 bp
* 97521 contig of 2154 bp in length
* 97621 gap of 100 bp
* 97854 contig of 2228 bp in length
* 99785 gap of 100 bp
* 99885 contig of 2020 bp in length
* 102113 gap of 100 bp
* 102112 contig of 2020 bp in length
* 102313 gap of 100 bp
* 104342 contig of 6358 bp in length
* 104343 gap of 100 bp
* 110710 contig of 2557 bp in length
* 110711 gap of 100 bp
* 113367 contig of 2028 bp in length
* 113467 gap of 100 bp
* 113468 contig of 6269 bp in length
* 115505 gap of 100 bp
* 115506 contig of 5122 bp in length
* 121875 gap of 100 bp
* 121876 contig of 5122 bp in length
* 127097 gap of 100 bp
* 127197 contig of 2428 bp in length

FEATURES
source

location/Qualifiers
1..129624
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9623"
/clone="PigB-211K18"
/clone_1fb="PigB"
1..4262
/note="assembly_fragment:00361
fragment_chain:1"
misc_feature
4363..6991
/note="assembly_fragment:00108
fragment_chain:1"
7092..9480
/note="assembly_fragment:00460
fragment_chain:1"
9581..11688
/note="assembly_fragment:00959
fragment_chain:1"
11789..14658
/note="assembly_fragment:00995
fragment_chain:2"
14759..18440
/note="assembly_fragment:00608
fragment_chain:2"
18541..20818
/note="assembly_fragment:01101
fragment_chain:2"
20919..22969
/note="assembly_fragment:00996
fragment_chain:2"
23070..27584
/note="assembly_fragment:00217
fragment_chain:3"
27685..33211
/note="assembly_fragment:00019
fragment_chain:3"
33312..35521
/note="assembly_fragment:00548
fragment_chain:3"
35622..38528
/note="assembly_fragment:00815
fragment_chain:4"
38629..42063
/note="assembly_fragment:00246
fragment_chain:4"
42164..44164
/note="assembly_fragment:00859
fragment_chain:5"
44265..47707
/note="assembly_fragment:00964
fragment_chain:5"
47808..50737
/note="assembly_fragment:01192
fragment_chain:6"
50838..57066
/note="assembly_fragment:00542
fragment_chain:6"
57167..59978
/note="assembly_fragment:01304
fragment_chain:7"
60079..62560
/note="assembly_fragment:00948
fragment_chain:7"
62661..64963
/note="assembly_fragment:00016"
65064..67191
/note="assembly_fragment:00065"
67292..72005
/note="assembly_fragment:00070"
72106..76595
/note="assembly_fragment:00136"
76696..79825

misc_feature
/note="assembly_fragment:00162"
79926..83540
/note="assembly_fragment:00367"
83641..86276
/note="assembly_fragment:00492"
86377..94620
/note="assembly_fragment:00545"
94721..97520
/note="assembly_fragment:00745"
97621..99784
/note="assembly_fragment:00801"

Alignment Scores:

Pred. No.:	4,38e-08	Length:	129624
Score:	383.00	Matches:	76
Percent Similarity:	57.66%	Conservative:	3
Best Local Similarity:	55.47%	Mismatches:	44
Query Match:	29.95%	Indels:	14
DB:	2	Gaps:	1

US-09-864-291-12 (1-235) x BX236540 (1-129624)

QY 96 Cyservallievallytygly-----102

DB 65506 TGCCCACTTGCTCTATGACCCCTACAGTAAATATGAGGCCAACAGCAATAT 65447

QY 103 ---Alapropoalaglytyglyalaproproproglytyglyalaproproalagly 121

DB 65446 AGAGCCCACTCCAGATATGAGAGCCCACTCCAGATATGAGAGCCCACTCCAGGA 65387

QY 122 Tyglyalaglaprovalaglyasngluglyproprovalaglytytytytytytytyty 141

DB 65386 TACGAGCCCACTCCAGATATGAGAGCCCACTCCAGATATGAGAGCCCACTCA 65327

QY 142 Argtyglyalaproproleuglytytytytytytytytytytytytytytytytyty 161

DB 65326 AGACTCGAAGCCCACTCCAGATATGAGAGCCCACTCCAGATATGAGAGCCCACT 65267

QY 162 Leuglyty 181

DB 65206 CCAAGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 65147

QY 202 serproalaglyserglyalargprogluglyserthralalaglnala 218

DB 65146 CCGCTCCAGATATGAGAGCTCCAGATATGAGAGCTCCAGATATGAGAGCTCC 65096

RESULT 13

AC092212 170781 bp DNA linear HTG 05-APR-2002

LOCUS AC092212 Trypanosoma brucei chromosome VIII clone RPI393-10J17, ***

DEFINITION SEQUENCING IN PROGRESS ***

AC092212 AC092212.31 GI:20043109

ACCESSION AC092212.31 GI:20043109

KEYWORDS HTG, HTGS, PHASE2.

SOURCE Trypanosoma brucei

ORGANISM Trypanosoma brucei

Bukariola; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

1 (bases 1 to 170781)

El-Sayed,N.M., Ghedin,B., Song,J., Larkin,C., Wanless,D., Jones,K.,

Peterson,J., Hou,L., Zhao,H., Mason,T., Millican,J., Pal,G., Van

Aken,S., Utecherback,T., Khalak,H.G., Gerard,C., Leech,V., Ullu,B.,

Melville,S., White,O., Adams,M.D., Donaldson,J.B. and Fraser,C.M.

Trypanosoma brucei GUTc10.1 RPI393-10J17 BAC genomic sequence

Unpublished

2 (bases 1 to 170781)

El-Sayed,N.M., Khalak,H. and Adams,M.D.

Direct Submission

JOURNAL Submitted (29-JUN-2001) The Institute for Genomic Research, 9712

REFERENCE 3 (bases 1 to 170781)
 Medical Center Dr, Rockville, MD 20850, USA
 AUTHORS El-Sayed,N.M., Khalak,H. and Adams,M.D.
 TITLE Direct Submission
 JOURNAL Submitted (05-APR-2002) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 COMMENT On Apr 5, 2002 this sequence version replaced gi:18425287.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 170781: contig of 170781 bp in length.
 * Location/Qualifiers
 1..170781
 /organism="Trypanosoma brucei"
 /mol_type="genomic DNA"
 /isolate="GUTat10.1"
 /db_xref="taxon:5691"
 /chromosome="VIII"
 /clone="RPC193-10J17"
 BASE COUNT 42610 a 38223 c 43377 g 46571 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.00102 Length: 170781
 Score: 289.00 Matches: 73
 Percent Similarity: 47.50% Conservative: 3
 Best Local Similarity: 45.62% Mismatches: 50
 Query Match: 22.60% Indels: 34
 Gaps: 11
 DB: 2
 US-09-864-291-12 (1-235) x AC092212 (1-170781)
 QY 87 GYAsmMeCythrProglMet-----ProCySerValIleVal 100
 Db 65879 GGGGGGAGATTGTTCCAGGCTTAATGTGCTGGCTGGGCTCCACAGATTAATGAAAA 65938
 QY 101 TrrglYAlaProProAlaGlyTrrglYAlaProProPro-----GlyTrrglY----- 116
 Db 65939 CAGCCCCCACCACCTGCGGATATGTCAACCCACACCTGCGGATATGTCAACC 65998
 QY 117 AlAProProAlaGlyTrrglYAlaGlyProValGlyAsnGlyProProValGlyTrr 136
 Db 65999 CCACCACTGCGGATATGTCAACCCCA-----CCACCTGCGGATAT 66043
 QY 137 ArgAlaSerProValaGlyTrrglYAlaProProleuGlyTrrglY-----AlaProPro 154
 Db 66044 GGTGAGCCCCA-----CCACCTGCGGATATGTCAACCCCAACCACT 66088
 QY 155 AlAglYTrglY-----AlAProProleuGlyTrrglY-----AlAProProleuGly 170
 Db 66089 GCCCGATATGTCAACCCCAACCACTGCGGATATGTCAACCCCAACCACTGCGCGA 66148
 QY 171 TrrglY-----ThrProProleuGlyTrrglY-----AlAProProleuGlyTrrglY 186
 Db 66149 TATGTGAGCCCCCAACCACTGCGGATATGTCAACCCCAACCACTGCGCGATATGT 66208
 QY 187 AlAProProAla-----GlyAsnGlyProProAlaGlyTrrglYArgAlaSer 202
 Db 66209 CAGCCCCCACCACCTGCGGATATGTCAACCCCAACCACTGCGGATATGTCAACCG 66268
 QY 203 ProAlaGlySerGlyAlaArgProGlnGlySerThrAlaAlaGlnAlaProGlnAsnGly 222
 Db 66269 CCGGTGTGCGCTTAAACAACCAACCGATGATATGTCTGAAGTCAAAAGTGGGAAGGAG 66328
 RESULT 14
 AB072784 2170 bp mRNA linear PRI 22-FEB-2003
 LOCUS

DEFINITION Macaca fascicularis testis cDNA clone:QcSA-20807, full insert
 sequence.
 ACCESSION AB072784
 VERSION 1
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Macaca fascicularis (crab-eating macaque)
 ORGANISM Macaca fascicularis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Butheria; Primates; Catarrhini; Cercopithecoidea;
 Cercopithecinae; Macaca.
 REFERENCE 1
 Otsada,N., Hida,M., Kusuda,J., Tanuma,R., Hirata,M., Suto,Y.,
 Hirai,M., Terao,K., Sugano,S. and Hashimoto,K.
 Cynomolgus monkey testicular cDNAs for discovery of novel human
 genes in the human genome sequence
 BMC Genomics 3 (1), 36 (2002)
 JOURNAL 12498619
 PUBMED 2 (bases 1 to 2170)
 REFERENCE Hashimoto,K., Otsada,N., Hida,M., Kusuda,J. and Sugano,S.
 Direct Submision
 JOURNAL Submitted (09-OCT-2001) Katsuyuki Hashimoto, National Institute of
 Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
 (E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genbank/
 Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
 Lab host: TOP10
 Vector: pME18S-FL3 (Acc.No. AB009864)
 R. Site1: DraIII (CACTGATG)
 R. Site2: DraIII (CACCAGTGT)
 Description: 1st strand cDNA was primed with an oligo(dT) primer
 [ATGAGCCTTTTCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
 using specific 5' and 3' primers and amplified by PCR. The PCR
 product was digested with SfiI and size selection was performed to
 exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
 into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
 the DraIII sites can be used to isolate the cDNA insert. Libraries
 were constructed by oligo-capping method
 (Sugano et al., University of Tokyo, Institute of Medical Science).
 Custom primer used for sequencing
 (5' end primer [CTTCTGCTTAAGACGTCG];
 3' end primer [CGACCTGACGCTCAACGAC])).
 Location/Qualifiers
 1..2170
 /organism="Macaca fascicularis"
 /mol_type="mRNA"
 /db_xref="taxon:9541"
 /clone="QcSA-20807"
 /sex="male"
 /tissue_type="testis"
 /clone_lib="macaque testis cDNA library QcSA"
 /dev_stage="adult"
 419..1171
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="BAB69753.1"
 /db_xref="GI:16041172"
 /translation="MFQNCSSGSLMSYDVELTNDKNEVBAFKTKGTYLTPT
 RYFLSKGDAMQSPMPYLMKDCIEIKQPVGANTIKGVAAAGGWSGASLYLT
 FTAGALIFGQNLGVASQASRGAPNAGLYSPYSGAVYVPPVAMQVYCPDPY
 YPPPPRYRPPMDGAMGAGVOPPPRYRGMPPVSGSDVPSPTAAAKAABAAS
 ATNPNRPNVNTPTSGPPPPRYPPEDKTY"
 BASE COUNT 464 a 663 c 544 g 499 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.00111 Length: 2170
 Score: 276.50 Matches: 90
 Percent Similarity: 38.49% Conservative: 22
 Best Local Similarity: 30.93% Mismatches: 76
 Query Match: 21.62% Indels: 103
 Gaps: 12
 DB: 9
 US-09-864-291-12 (1-235) x AB072784 (1-2170)

```

Oy 1 MetProPheapLeuMetThrasnLeuThralValGluInProValPheAlaAlaasnphe 20
Db 599 ATGCGGTTTATCTCATGAAAGACGTGATGACGAGCCCGTATTGGTGCACAACTAC 658
Oy 21 ILeysglYThrlIeglnAlaAlaProTyrglyYTPrglyYglAlaAlaThrPheYs 40
Db 659 ATCAAGGACACAGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 718
Oy 41 LeuValPheArGanGlyAspAlaIleGluPheAlaGlnLeuMetValAlaAlaIAser 60
Db 719 TTGACTTTACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 778
Oy 61 AlaValAlaArGlyPheProLeuArGhrLeuAsnArTPheSerSerMetGlyYle 80
Db 779 CAAGCTTCAGAGGT-----GAAGCCCAAGTACAGACC 811
Oy 81 TyrValIleThrlGlyGluYAsnMetCysThrProGlnMetProCysSerValIleVal 100
Db 812 TATGGCTACTCT-----TACATGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 847
Oy 101 TyrglyAlaProProAlaGlyTyrglyAlaPro---ProProGlyTyrglyAlaProPro 119
Db 848 TATCCCCCGGAGTGCACATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 907
Oy 120 AlaGlyTyrglyAlaGlnProValGlyAlaGlnGlyProProValGlyTyrglyAlaSer 139
Db 908 CCC-----CCGCTGAGTTCATTCAGAGACC 934
Oy 140 ProValArGlyTyrglyAla-----ProProLeuGlyTyrglyAlaPro 153
Db 935 CCCATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 994
Oy 153 -----CCGCTGAGTTCATTCAGAGACC 153
Db 995 ATGAACTCCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1054
Oy 154 -----Pro 154
Db 1055 GCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1114
Oy 155 AlaGlyTyrglyAlaProProLeuGlyTyrgly----- 165
Db 1115 ACGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1174
Oy 166 -----AlaProProLeuGly---TyrglyThrProProLeuGlyTyrgly 179
Db 1175 CTCCTGCTCCCTGCTCCCAACCTTATCTGTAACCTTATCTGTAACCTTATCTGTAAC 1231
Oy 180 AlaProProLeuGlyTyrglyAlaProProAlaGlyAlaGlnGlyProProAlaGlyTy 199
Db 1232 GCACGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1282
Oy 200 ArgAlaSerProAlaGlySer-----GlyAlaArGProGlnGlu----- 212
Db 1283 AAACATTATCAAGAACTATGATGAGACATTAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1342
Oy 213 -----SerThralAlaGlnAlaProGlu 220
Db 1343 CGCCAGCTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1375

```

```

RESULT 15
AC134957 166045 bp DNA linear HTG 23-NOV-2002
LOCUS AC134957
DEFINITION Tetradon nigrorivialis clone GSTMB-4505, WORKING DRAFT SEQUENCE, 3
ordered pieces.
ACCESSION AC134957
VERSION AC134957.2 GI:25188937
KEYWORDS HTG; HTGS; PHASE2; HTGS; DRAFT.
SOURCE Tetradon nigrorivialis
ORGANISM Tetradon nigrorivialis
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Acanthomorpha; Acanthopterygii; Percomorpha; Tetradontiformes;
Tetradontoidae; Tetradontidae; Tetradon.
1 (bases 1 to 166045)
Ahter, N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bonfield, G.G., Brinkley, C., Brooks, S.,
Carliaga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,
Haghighi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, B.,
Laric, P., Lee-Lin, S.-O., Legaspi, R., Meduro, O.L., Meduro, V.B.,
Margiles, B.H., Mastello, C., Marker, B., McDowell, J.,
Paguitigan, C., Pearson, R., Portnoy, M.B., Prasad, A.,
Reddy-Dugue, N., Schandler, K., Schneider, M.G., Sison, C.,
Stentilpop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,
Weberby, K.D., Wiggins, L., Young, A. and Green, B.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 166045)
Green, B.D.
Direct Submission
Submitted (03-OCT-2002) NIH Intramural Sequencing Center, 8717
GroveMont Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 166045)
Green, B.D.
Direct Submission
Submitted (23-NOV-2002) NIH Intramural Sequencing Center, 8717
GroveMont Circle, Gaithersburg, MD 20877, USA
On Nov 23, 2002 this sequence version replaced gi:23477837.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.zoo@nih.gov
----- Project Information
Center project name: dgd
Center clone name: 045005

The sequence data in this record represents an 'enhanced' version of a phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165749 bases at least Q40
Consensus quality: 165822 bases at least Q30
Consensus quality: 165836 bases at least Q20
Insert size: 175000; agarose-fp
Insert size: 165845; sum-of-contigs
Quality coverage: 13.18x in Q20 bases; sum-of-contigs
Quality coverage: 13.91x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1
* 44384: contig of 44384 bp in length
* 44385
* 44484: gap of unknown length
* 44485
* 55090: contig of 10606 bp in length
* 55091
* 55190: gap of unknown length
* 55191
* 166045: contig of 110855 bp in length.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: December 16, 2003, 11:00:59 ; Search time 196.834 Seconds

(without alignments)
3222.858 Million cell updates/sec

Title: US-09-864-291-12

Perfect score: 1279
Sequence: 1 MPDMLTNTLTVQPVFANF.....AQAPENASLPASSSQVHS 235

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1345719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ .p2n.model -DRV=xlp
-O=/csg12.1/USPTO.spool.p/US09864291/runat.15122003.160858.23296/app.query.fasta_1.846
-DB=N Geneseq.19Jun03 -OPMT=fastcap -SUFFIX=ring -MIMATCH=0.1 -LOOPTC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62 -TRANS=human40.cdt
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09864291.OCEN.1.1.511.OTUNAT.15122003.160858.23296 -NCP=6 -ICPU=3
-NO MAMP -LARGEQUERY -NEG SCORES=0 -NAIT -DSBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -MAIN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq.19Jun03.*
1: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1279	100.0	1001	24	AAS20602	DNA encoding human
2	826	64.6	467	22	ABL01229	Human reproductive
3	826	64.6	467	23	ABL96688	Human testicular a
4	777	60.8	1413	24	AAS20601	DNA encoding bovin
5	765	59.8	220895	24	ABK84798	Human cDNA differe
6	761	59.5	436	22	ABA67850	Human foetal liver
7	761	59.5	436	22	AAK42003	Human bone marrow
8	761	59.5	436	22	AAI48070	Probe #16755 used
9	761	59.5	436	24	ABSI6034	Human genome-deriv
10	705	55.1	7099	22	AAI04882	Human reproductive
11	705	55.1	7099	23	ABJ79776	Human testicular a
12	342	26.7	894	23	AAS70582	DNA encoding novel
13	261.5	20.4	1885	24	ABJ70430	Human bone remodel
14	258	20.2	1915	21	AAJ22363	Human secreted pro
15	258	20.2	1915	25	ABJ73614	Secreted protein-e
16	258	20.2	1915	25	ABJ76841	Human secreted pro
17	258	20.2	1915	25	ABJ67208	Human secreted pro
18	252.5	19.7	1109	23	ABJ17151	Drosophila melanog
19	245	19.2	471	22	ABA55346	Human foetal liver
20	245	19.2	471	22	AAK28957	Human bone marrow
21	245	19.2	471	22	AAI34906	Probe #3592 used t
22	245	19.2	471	24	ABJ03491	Human genome-deriv
23	220.5	17.2	4881	19	AAV63438	Feline herpesvirus
24	215	16.8	2614	23	ABL28430	Drosophila melanog
25	215	16.8	2614	23	ABL05187	Drosophila melanog
26	215	16.8	9516	23	ABL05186	Drosophila melanog
27	209.5	16.4	756	17	AAV16766	Drosophila-like poly
28	206.5	16.1	756	14	AAQ43032	Collagen-like poly
29	204.5	16.0	4329	23	ABL28034	Drosophila melanog
30	202.5	15.8	1863	23	ABL28035	Drosophila melanog
31	197.5	15.4	4323	23	ABL17150	Drosophila melanog
32	195	15.2	432	17	AAV16765	Collagen-like poly
33	194	15.1	468	23	ABL28431	Drosophila melanog
34	192.5	15.1	2833	23	ABL28433	Drosophila melanog
35	192.5	15.1	6906	23	ABL29392	Drosophila melanog
36	192	15.0	321	22	AAI01355	Human reproductive
37	192	15.0	321	23	ABL96808	Human testicular a
38	192	15.0	4403765	22	AAI99683	Mycobacterium tube
39	192	15.0	4411529	22	AAI99682	Mycobacterium tube
40	191	14.9	758	22	AAI23504	Human breast cance
41	189.5	14.8	3206	23	ABL16637	Drosophila melanog
42	189.5	14.8	5595	23	ABL16636	Drosophila melanog
43	188.5	14.7	432	14	AAQ43031	Collagen-like poly
44	186	14.5	2144	18	AAV69165	Trypanosoma cruzi
45	186	14.5	2144	20	AAK81754	DNA encoding a T.

ALIGNMENTS

RESULT 1
AAS20602
ID AAS20602 standard; cDNA, 1001 BP.
AC AAS20602;
XX
XX 09-APR-2002 (first entry)
XX
XX DNA encoding human testicular WW domain binding protein (htwbp).
XX
XX Testicular WW domain binding protein; WBPI; perinuclear theca 32;
XX P332; contraceptive; fertility; oocyte activation; vaccine;
XX globozoosperm; spermatogenesis; spermatocyst; tyrosine kinase; c-Yes;
XX immunocarcinogenic; human; gene; ss.
XX
XX Homo sapiens.
XX

```

FH Key Location/Qualifiers
FT primer_bind 1..18
      /tag= a
      /note= "Binds forward primer AAS20605"
FT CDS 1..708
      /tag= b
      /product= "htMBP"
FT primer_bind /note= "Human testicular WW domain binding protein"
FT complement (984..1001)
FT /tag= c
FT /note= "Binds reverse primer AAS20606"
XX
XX MO200190185-A2.
XX
XX 29-NOV-2001.
XX
XX 25-MAY-2001; 2001WO-CA00738.
XX
XX 25-MAY-2000; 2000CA-2307128.
XX 25-MAY-2000; 2000US-206979P.
XX
XX (TOOH ) UNIV QUEENS KINGSTON.
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Oko R, Sutovsky P;
XX
XX WPI: 2002-097644/13.
XX DR P-PSDB; AAU74610.
XX
XX Isolated perinuclear theca 32 polypeptide that interacts with activated
XX tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing
XX diminished fertility and abnormal spermiogenesis and for providing
XX contraception -
XX
XX Claim 62; Fig 4B; 1033p; English.
XX
XX The invention describes an isolated perinuclear theca 32 (PT32)
XX polypeptide (1) which interacts with tyrosine kinase c-Yes. (1) is
XX useful for: enhancing fertility in a mammal; treating globozoospermy, by
XX expressing (1) in spermatozoa; inhibiting fertilisation, by introducing
XX (1) or its antigenic fragment into a mammal to elicit an immune
XX response; enhancing the ability of round spermatids to activate oocytes;
XX treating or diagnosing diminished fertility and abnormal spermiogenesis;
XX in providing contraception; identifying contraceptive and
XX fertility-enhancing agents. The polynucleotide is useful for producing
XX (1) by recombinant techniques, as vaccine, as diagnostic reagents, and
XX for chromosome identification. An antibody against (1) is useful in
XX immunological assays, in immun contraceptive methods, to identify cells
XX expressing (1), and to purify (1) by affinity chromatography. A
XX transgenic animal is useful as an animal model for studying human
XX fertility and reproductive biology, and for screening compounds to
XX identify modulators of oocyte activation. The use of (1) prevents the
XX entry of components which are detrimental to embryonic development into
XX the oocyte during oocyte activation with crude sperm extract and avoids
XX the propagation of viruses such as HIV (human immunodeficiency virus) and
XX SIV (Sialian immunodeficiency virus) carried in the sperm. This sequence
XX encodes the human testicular WW domain binding protein (htMBP), described
XX in the method of the invention.
XX
XX SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 2.58e-58 Length: 1001
XX Score: 1279.00 Matches: 235
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Gaps: 0
XX DB: 24
XX
XX US-09-864-291-12 (1-235) x AAS20602 (1-1001)
XX
XX 1 MetProPheAspLeuMetThrAsnLeuThrValGluGlnProValPheAlaIaenPhe 20
XX

```

Db	1	ATGCCATTGATCTGATGACGAACCTCAGCTGTGAAACAACAGATTGTCGCAAACTTC	60
Qy	21	IIeLyeGLYThrTIIeGlnAlaIaProTyrTgIyGItYrTgIuGIyGlnAlaThrPheLys	40
Db	61	ATTAAAGGAACATTATCAGGACAGCTCCATATGTGTGCTGGGAAGACAAGACTTATTTAA	120
Qy	41	LeuValPheArgAsnGIyAspAlaIIeGluPheAlaGlnIleuMetValIyValAlaSer	60
Db	121	TTTGCTCTCAGAAATGAGATGCCATGTAAATTGGCCAGTGTGATGTGAAGCTGCTCT	180
Qy	61	AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle	80
Db	181	GCCTGTGCCGAGGANTTCCACTTAGAACCTTAAATGATCTGTTCAGCTCATATGGAA	240
Qy	81	TyrValIleThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleVal	100
Db	241	TATGTAAATTACTGGGGAGAGAAATATGGCACTCCACAGATGCTTGTTCAGTAAATGTC	300
Qy	101	TyrGlyAlaProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaProProAla	120
Db	301	TATGGGCCCCCAGCTGAGAAATATGAGCCCACTCCCGAATGAGAGCCCACTGCA	360
Qy	121	GlyTyrGlyAlaGlnProValGlyIyAsnGluGlyProProValGlyTyrArgAlaSerPro	140
Db	361	GGATATGAGAGCCCAACCGTAGAAATGAAAGCCCGCGCTGTGGATACAGAGCTCACT	420
Qy	141	ValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaPro	160
Db	421	GTGCGATATGAGCCCCCACTCTGTGATATCGAAGCCCACTGCGAGATATGAGAGCCCA	480
Qy	161	ProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAla	180
Db	481	CCCTTAGAATATGAGAGCCCACTCTGTGATATGAGAGCCCACTCTGTGATATGAGAGCC	540
Qy	181	ProProLeuGlyTyrGlyAlaProProAlaGlyIyAsnGluGlyProProAlaGlyTyrArg	200
Db	541	CCACCTCTCGAATATGAGAGCCCACTCTGAGAAATGAAAGCCCGCGCGGATACAGA	600
Qy	201	AlaSerProAlaGlySerGlyAlaArgProGlnGlnSerThrAlaAlaGlnAlaProGlu	220
Db	601	GGCTCACCTGCTGATCAGAGAGCCAGGCTCAGAAATCTACAGCAGGCCAGGCTCTGAA	660
Qy	221	AsnGluAlaSerIleuProSerAlaSerSerSerGlnValHiser 235	
Db	661	AACGAGGCTTCTCTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	705
RESULT 2			
XX	AA101229	standard, cDNA; 467 bp.	
XX	AA101229;		
DT	21-NOV-2001	(first entry)	
XX			
DE		Human reproductive system related antigen cDNA SEQ ID NO: 1230.	
KW		Human; reproductive system related antigen; reproductive system disorder	
KM		cancer; gene therapy; ss.	
XX			
OS		Homo sapiens.	
XX			
PN		MO200155320-A2.	
PD			
XX		02-AUG-2001.	
PF			
XX		17-JAN-2001; 2001MO-US01339.	
PR		31-JAN-2000; 2000US-0179065.	
PR		04-FEB-2000; 2000US-0180628.	
PR		24-FEB-2000; 2000US-0184664.	
PR		02-MAR-2000; 2000US-0186350.	
PR		16-MAR-2000; 2000US-0189874.	
PR		17-MAR-2000; 2000US-0190076.	

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 26-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225269.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX MPI; 2001-465570/50.
XX P-PSDB; AAM95259.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 1; SEQ ID NO 1230; 1297bp + Sequence listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC

CC including cancer. The present sequence is a coding sequence of the
invention.

XX Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;

Alignment Scores:

pred. No.: 3,68e-35 Length: 467
Score: 826.00 Matches: 149
Percent Similarity: 96.77% Conservative: 1
Best Local Similarity: 96.13% Mismatches: 4
Query Match: 64.58% Indels: 1
DB: 22 Gaps: 0

US-09-864-291-12 (1-235) x AAL01229 (1-467)

QY 44 ArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIysAlaSerAlaValAla 63
DB 3 AGAAATGAGAGTGCCTTGAATTTGGCCAGTTGATGTAAGAACTGCTGCTGCC 62
QY 64 ArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerMetGlyIleValIle 83
DB 63 CGAGGATTTCCACTTGAACCTTAATGACCTGGTTCAGCTCTAAGGAATTTATGTAAT 122
QY 84 ThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleValIleValIle 103
DB 123 ACTGGGGAAGGAAATATGTCACCTCCACAGATGCTTGTCAGTATTGCTAAGAGCC 182
QY 104 ProProAlaGlyIleGlyAlaProProProGlyIleGlyAlaProProAlaGlyIleGly 123
DB 183 CCACCTGACGAGATATGAGGCCCCACCTCCGAGATACGAGGCCCCACCTGAGATATGGA 242
QY 124 AlaGlnProValGlyAsnGluGlyProProValGlyIleGlyAlaSerProValArgIle 143
DB 243 GCCCAACCCGTAGGAATATGAGGCCCGCTGTGGATATACAGAGCTCAGCTGTGCATAT 302
QY 144 GlyAlaProProLeuGlyIleGlyAlaProProAlaGlyIleGlyAlaProProLeuGly 163
DB 303 GGAGCCCCACCTCTGATACGAGGCCCCACCTGACGAGATATGAGGCCCCACCTTAGGA 362
QY 164 TyrGlyAlaProProLeuGlyIleGlyAlaProProLeuGlyIleGlyAlaProProLeu 183
DB 363 TATGAGGCCCAACCTTTGATATGAGCCCACTCTGAGATATGAGGCCCACTTNT 422
QY 183 uGlyIleGlyAlaProProAlaGlyAsnGluGlyProProAla 197
DB 423 CGGATATGAGGCCCACTCTGAGATATGAGGCCCGCTTCG 465

RESULT 3

ABL96688 standard; cDNA; 467 BP.

XX ABL96688;

XX 21-JUN-2002 (first entry)

XX Human testicular antigen encoding cDNA SEQ ID NO: 356.

XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX reproductive system disorder; urinary system disorder; gene therapy;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disease; infection; cytostatic; gene; ss.

OS Homo sapiens.

XX WO200155317-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01329.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227189.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229513.
PR 05-SEP-2000; 2000US-0229519.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232168.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235634.
PR 27-SEP-2000; 2000US-0235636.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236328.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240950.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483232/52.
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating testicular cancer -
 XX
 PS Claim 1; SEQ ID NO 356; 766pp; English.
 XX
 CC The present invention provides the protein and coding sequences of 973

CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a cDNA of the
 CC invention.
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
 Alignment Scores:
 Pred. No.: 3,68e-35 Length: 467
 Score: 826.00 Matches: 149
 Percent Similarity: 96.77% Conservative: 1
 Best Local Similarity: 96.13% Mismatches: 4
 Query Match: 64,588 Indels: 1
 DB: 23 Gaps: 0
 US-09-864-291-12 (1-235) x ABL96688 (1-467)
 QY 44 ArgAaenglyAaspAlallegluPhaAlagInleuMetVallyAlaAlaSerAlaValAla 63
 DB 3 AGAAATGAGAGTGCCTTAATGATTTGCCAGTTGATGTAAGAGCTGCTGCTGCC 62
 QY 64 ArgGlyPheProleuArgThrIleuSnaApITpPheSerSerMetGlylleTyrrValIle 83
 DB 63 CAGGATTCACACTTAAGACCTTAATGACTGATTCAGCTTAAGGAAATTAATTAAT 122
 QY 84 ThrGlygluglyAaMetCyethrProglMetProCySerValIleValTyrrGlyAla 103
 DB 123 ACTGGGAAAGGGAATATGTCACCTCAGAGATGCTTTCATTAATTCATAGAGCC 182
 QY 104 ProProAlaGlyTyrrGlyAlaProProProGlyTyrrGlyAlaProProAlaGlyTyrrGly 123
 DB 183 CCACTGCAGATATGAGAGCCCACTCCGATTCAGAGCCCACTGCAGATATGGA 242
 QY 124 AlaGlnProValGlyAaenglygluProProValGlyTyrrArgAlaSerProValArgTyrr 143
 DB 243 GCCCAACCCGTAGAAATAGAGCCCGCTGAGATAGAGAGCCCACTGTCATAT 302
 QY 144 GlyAlaProProleuGlyTyrrGlyAlaProProAlaGlyTyrrGlyAlaProProleuGly 163
 DB 303 GAGAGCCCACTCTTGATAGAGAGCCCACTGCAGATATGAGAGCCCACTCTAGGA 362
 QY 164 TyrrGlyAlaPro-ProleuGlyTyrrGlyThrProProleuGlyTyrrGlyAlaProProle 183
 DB 363 TATGAGAGCCCAACTTGTGATATGAGAGCCCACTCTGATATGAGAGCCCACTT 422
 QY 183 uGlyTyrrGlyAlaProProAlaGlyAaenglygluProProAla 197
 DB 423 CGGATATGAGAGCCCACTGCAGAGAAATGAGCCCGCTTGG 465
 RESULT 4
 AAS20601
 ID AAS20601 standard; cDNA; 1413 BP.
 AC AAS20601;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE DNA encoding bovine perinuclear theca 32 (PT32).
 XX
 KW Testicular WW domain binding protein; hTWBP; perinuclear theca 32;
 KW PT32; contraceptive; fertility; oocyte activation; vaccine;
 KW globzoospermy; spermiogenesis; spermatocoe; tyrosine kinase; c-Yes;
 KW Immunoc contraceptive; bovine; gene; ss.
 OS Bos sp.
 XX
 XX Key Location/Qualifiers
 FT primer_bind 30..50
 FT /note= "primer binding site for cDNA isolation. The
 FT sequence differs from that of the forward primer

```

FT CDS          36..977      given in AAS20603"
FT              /tag= b
FT              /product= "PTJ2"
FT protein_bind /note= "Perinuclear theca 32"
FT              978..1001
FT              /*tag= c
FT              /note= "Primer binding site for cDNA isolation. The
FT                      sequence differs from that of the reverse primer
FT                      given in AAS20604"
FT
FT WO200190185-AZ.
FT
FT 29-NOV-2001.
FT
FT 25-MAY-2001; 2001WO-CA00738.
FT
FT 25-MAY-2000; 2000CA-2307128.
FT 25-MAY-2000; 2000US-206979P.
FT
FT (TOOH ) UNIV QUEBENS KINGSTON.
FT (UYOR-) UNIV OREGON HEALTH SCI.
FT
FT Oko R, Sutovsky P;
FT
FT WPI; 2002-097644/13.
FT P-PsDB; AAU74604.
FT
FT Isolated perinuclear theca 32 polypeptide that interacts with activated
FT tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing
FT diminished fertility and abnormal spermiogenesis and for providing
FT contraception -
FT
XX PS Claim 10; Fig 2A-B; 103pp; English.
XX
CC The invention describes an isolated perinuclear theca 32 (PTJ2)
CC polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is
CC useful for: enhancing fertility in a mammal; treating globozoosperm, by
CC expressing (I) in spermatozoa; inhibiting fertilisation, by introducing
CC (I) or its antigenic fragment into a mammal to elicit an immune
CC response; enhancing the ability of round spermatids to activate oocytes;
CC treating or diagnosing diminished fertility and abnormal spermiogenesis;
CC in providing contraception; identifying contraceptive and
CC fertility-enhancing agents. The polynucleotide is useful for producing
CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
CC for chromosome identification. An antibody against (I) is useful in
CC immunological assays, in immuncontraceptive methods, to identify cells
CC expressing (I), and to purify (I) by affinity chromatography. A
CC transgenic animal is useful as an animal model for studying human
CC fertility and reproductive biology, and for screening compounds to
CC identify modulators of oocyte activation. The use of (I) prevents the
CC entry of components which are detrimental to embryonic development into
CC the oocyte during oocyte activation with crude sperm extract and avoids
CC the propagation of viruses such as HIV (human immunodeficiency virus) and
CC SIV (Simian Immunodeficiency Virus) carried in the sperm. This sequence
CC encodes the bovine perinuclear theca 32 (PTJ2), described in the method
CC of the invention.
CC
SQ Sequence 1413 BP; 377 A; 363 C; 369 G; 304 T; 0 other;

```

[illegible]

```

XX 03-OCT-2001; 2001WO-US30821.
XX
XX 03-OCT-2000; 2000US-237189P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Beazer-Barclay Y, Weiseman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression
XX of genes associated with granulocyte activation, which serves as
XX diagnostic markers that is useful for monitoring disease states and
XX drug toxicity.
XX
XX Claim 1; SEQ ID NO 1369; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (gs) identified by
XX DNA chip analysis as given in the specification, and comparing
XX the expression level to an expression level in an unactivated
XX GC, where differential expression of Gs is indicative of GCA.
XX Also included are modulating (M2) GA by contacting GC with an agent
XX that alters the expression of at least one gene in Gs; (2) screening (M3)
XX for an agent capable of modulating GCA or an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease using the
XX gene expression profile; (3) detecting (M4) an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease, by detecting the
XX level of expression of the gene is indicative of inflammation;
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,
XX an allergic response in a subject, exposure of a subject to a pathogen
XX or sterile inflammatory disease, by contacting a tissue having
XX inflammation with an agent that modulates the expression of gene(s)
XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
XX modulating GA; M3 is useful for screening an agent capable of modulating
XX GCA preferably in an inflammation in a tissue; M4 is useful for
XX detecting an inflammation (especially chronic) in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
XX inflammatory injury, ARDS, adult respiratory distress syndrome,
XX inflammatory bowel disease, Crohn's disease, ulcerative colitis,
XX periodontal disease; also bacterial infection, viral infection,
XX parasitic infection, protozoal infection, fungal infection, and MS is
XX useful for treating one of the above conditions. The present
XX sequence represents a gene differentially expressed in granulocytes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 220895 BP; 53760 A; 54808 C; 55831 G; 56496 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 9, 74e-30 Length: 220895
XX Score: 765.00 Matches: 138
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 59.81% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-864-291-12 (1-235) x ABA6798 (1-220895)
XX
XX 98 ValilevalTYTGlyAlaProProAlaGlyTYTGlyAlaProProGlyTYTGlyAla 117
XX |||||
XX DB 170005 GTTATGCTATGAGCCCACTGCAAGATATGAGCCCACTCCGGAATACGAGACC 170064
XX |||||
XX QY 118 ProProAlaGlyTYTGlyAlaGlnProValGlyAangGlyGlyProProValGlyTYTGly 137
XX |||||

```

```

DB 170065 CCACCTGCAAGATATGAGCCCAACCGTAGAATAAGGCCCGCTGTGAGATACAGA 170124
QY 138 AlaseProValaGlyTYTGlyAlaProProLeuGlyTYTGlyAlaProProAlaGlyTYT 157
XX |||||
DB 170125 GCTTCACTGTGCAATATGAGCCCACTCTTGGAATACGAGCCCACTGCAAGATAT 170184
QY 158 G1YAlaProProLeuGlyTYTGlyAlaProProLeuGlyTYTGlyThProProLeuGly 177
XX |||||
DB 170185 GAGCCCCACCTCTAGATATGAGCCCACTCTTGATATGAAACCCCACTCTGGA 170244
QY 178 TyTGlyAlaProProLeuGlyTYTGlyAlaProProAlaGlyAangGlyGlyProProAla 197
XX |||||
DB 170245 TATGAGCCCACTCTGCAATATGAGCCCACTGCAAGAATGAGCCCGCTGCG 170304
QY 198 G1YTYTATGAlaSerProAlaGlySerGlyAlaAlaProGlnGlySerThrAlaAlaGln 217
XX |||||
DB 170305 GGAATACGAGGCTCACTGCTGATACAGAGCCGAGGCTCAGGAATATACAGAGCCAG 170364
QY 218 AlAProGluAangGlyAlaSerLeuProSerAlaSerSerGlnValHisSer 235
XX |||||
DB 170365 GCTCTGAAAACGAGGCTTCTTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 170418
XX |||||
RESULT 6
ABA67850
ID ABA67850 standard; DNA; 436 BP.
XX
XX ABA67850;
AC
XX
XX 01-FEB-2002 (first entry)
DE
XX Human foetal liver single exon nucleic acid probe #16155.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX MO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human foetal liver -
XX
XX Claim 4; SEQ ID NO 16155; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
XX

```

Alignment Scores:

Pred. No.: 7,99e-32 Length: 436
 Score: 761.00 Matches: 137
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 59.50% Indels: 0
 DB: 22 Gaps: 0

US-09-864-291-12 (1-235) x ABA67850 (1-436)

QY 99 ILevalTYrGlyAlaProProAlaGlyTYrGlyAlaProProGlyTYrGlyAlaPro 118
 DB 3 ATTGCTATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAGAGCCCA 62
 QY 119 ProAlaGlyTYrGlyAlaGlnProValGlyAaGlnGlyProProValGlyTYrAaGAla 138
 DB 63 CCGGAGATATGAGAGCCCACTGAGAAATGAAGCCCGCTGGGATACAGAGCC 122
 QY 139 SerProValArGlyTYrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyTYrGly 158
 DB 123 TCACTGTGCGATATGAGAGCCCACTTGTGATACGAGAGCCCACTGAGATATGGA 182
 QY 159 AlaProProLeuGlyTYrGlyAlaProProLeuGlyTYrGlyThrProProLeuGlyTYr 178
 DB 183 GCCCACTTATGAGATATGAGAGCCCACTTGTGATATGAGAGCCCACTCCGGATAT 242
 QY 179 GlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyAaGlnGlyProProAlaGly 198
 DB 243 GGAGAGCCCACTTGTGATATGAGAGCCCACTGAGAAATGAAGCCCGCTGGGAG 302
 QY 199 TyrArGAlaSerProAlaGlySerGlyAlaArGProGlnGlySerThrAlaAlaGlnAla 218
 DB 303 TACAGAGCCCTCACTGTGATATGAGAGCCCACTTGTGATATGAGAGCCCACTGAGAT 362
 QY 219 ProGluAaGlnAlaSerLeuProSerAlaSerSerSerGlnValHisSer 235
 DB 363 CCGAAGAGAGAGGCTTCTTCCCTGCTGCTCCTTCAAGTCCATTCT 413

RESULT 7

AAK42003 ID AAK42003 standard; DNA; 436 BP.

AC AAK42003;

XX 06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 16560.

KM Human; bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.

OS Homo sapiens.

PN W0200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX Example 4; SEQ ID NO: 16560; 658bp + Sequence listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX

SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Alignment Scores:

Pred. No.: 7,99e-32 Length: 436
 Score: 761.00 Matches: 137
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 59.50% Indels: 0
 DB: 22 Gaps: 0

US-09-864-291-12 (1-235) x AAK42003 (1-436)

QY 99 ILevalTYrGlyAlaProProAlaGlyTYrGlyAlaProProGlyTYrGlyAlaPro 118
 DB 3 ATTGCTATGAGAGCCCACTGAGATATGAGAGCCCACTCCGGATACGAGAGCCCA 62
 QY 119 ProAlaGlyTYrGlyAlaGlnProValGlyAaGlnGlyProProValGlyTYrAaGAla 138
 DB 63 CCGGAGATATGAGAGCCCACTGAGAAATGAAGCCCGCTGGGATACAGAGCC 122
 QY 139 SerProValArGlyTYrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyTYrGly 158
 DB 123 TCACTGTGCGATATGAGAGCCCACTTGTGATACGAGAGCCCACTGAGATATGGA 182
 QY 159 AlaProProLeuGlyTYrGlyAlaProProLeuGlyTYrGlyThrProProLeuGlyTYr 178
 DB 183 GCCCACTTATGAGATATGAGAGCCCACTTGTGATATGAGAGCCCACTCCGGATAT 242
 QY 179 GlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyAaGlnGlyProProAlaGly 198
 DB 243 GGAGAGCCCACTTGTGATATGAGAGCCCACTGAGAAATGAAGCCCGCTGGGAG 302
 QY 199 TyrArGAlaSerProAlaGlySerGlyAlaArGProGlnGlySerThrAlaAlaGlnAla 218
 DB 303 TACAGAGCCCTCACTGTGATATGAGAGCCCACTTGTGATATGAGAGCCCACTGAGAT 362
 QY 219 ProGluAaGlnAlaSerLeuProSerAlaSerSerSerGlnValHisSer 235
 DB 363 CCGAAGAGAGAGGCTTCTTCCCTGCTGCTCCTTCAAGTCCATTCT 413

RESULT 8

AAI48070 ID AAI48070 standard; DNA; 436 BP.

AC AAI48070;

XX 17-OCT-2001 (first entry)

DE Probe #16756 used to measure gene expression in human placenta sample.

KM Probe: microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

OS Homo sapiens.

PN W0200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 16-MAR-2000; 2000US-0189874.

[illegible]

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0255719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
XX
XX Disclosure; SEQ ID NO 7570; 1297bp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC

CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
Alignment Scores:
Pred. No.: 6,71e-28 Length: 7099
Score: 705.00 Matches: 125
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 55.12% Indels: 0
DB: 22 Gaps: 0
US-09-864-291-12 (1-235) x AAL04882 (1-7099)
QY 98 ValIleValIYrGlyAlaProProAlaGlyTYrGlyAlaProProGlyTYrGlyAla 117
DB 6725 GTATTGTCTATGAGGCCCACTGAGATATGAGACCCCACTCCGATACGAGGCC 6784
QY 118 ProProAlaGlyTYrGlyAlaGlnProValGlyYasnGluGlyProProValGlyTYrArg 137
DB 6785 CCACTTCAGATATGAGCCCAACCCGTAGAAATGAAAGCCCGCTGTGGATACGA 6844
QY 138 AlaSerProValAlaGlyTYrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyTYr 157
DB 6845 GCCTCACCTGAGCATATGAGCCCACTTGTGATACGAGGCCCACTGCGAGATAT 6904
QY 158 GlyAlaProProLeuGlyTYrGlyAlaProProLeuGlyTYrGlyTYrProProLeuGly 177
DB 6905 GGAGCCCACTTATGAGATATGAGCCCACTTGTGATATGAGACCCCACTCTCGGA 6964
QY 178 TyrGlyAlaProProLeuGlyTYrGlyAlaProProAlaGlyYasnGluGlyProProAla 197
DB 6965 TATGAGCCCACTCTCGATATGAGCCCACTGCAAGAAATGAAAGCCCGCTGCG 7024
QY 198 GlyTYrArgAlaSerProAlaGlySerGlyAlaArgProGlnGlnSerThrAlaAlaGln 217
DB 7025 GGATACAGAGCCTCACCTGCTGATCAGAGCCGAGCCTCAGGAATCTACAGCAGCCGAG 7084
QY 218 AlaProGluGlnGlu 222
DB 7085 GCTCCTGAAAAAGAG 7099
RESULT 11
ABL97776
ID ABL97776 standard; DNA; 7099 BP.
XX
XX ABL97776;
AC
XX
XX 21-JUN-2002 (first entry)
DT
XX
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2428.
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX reproductive system disorder; urinary system disorder; gene therapy;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disease; infection; cytostatic; gene; ds.
XX
XX Homo sapiens.
OS
XX
XX
XX
XX MO200155317-A2.
XX
XX 02-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01329.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.

CC especially testicular cancers. The present sequence is a DNA encoding a
CC protein fragment of the invention.

XX Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;

Alignment Scores:

Pred. No.:	6,71e-28	Length:	7099
Score:	705.00	Matches:	125
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	55.12%	Indels:	0
DB:	23	Gaps:	0

US-09-864-291-12 (1-235) x ABL97776 (1-7099)

```
OY 98 ValIleValIleTYRGIYAlaProProAlaGIYTYRGIYAlaProProGIYTYRGIYAla 117
DB 6725 GTTATGTCTATGAGCCCACTTCAGATATGAGCCCACTCCCGATACCGAGCC 6784
OY 118 ProProAlaGIYTYRGIYAlaGIaProProAlaGIYTYRGIYAlaProProAlaGIYTYR 137
DB 6785 CCACTGAGATATGAGCCCACTTCAGATATGAGCCCACTCCCGATACCGAGCC 6844
OY 138 AlaSerProValArgTYRGIYAlaProProLeuGIYTYRGIYAlaProProAlaGIYTYR 157
DB 6845 GCTTACCTGTGCGATATGAGCCCACTTCGATATGAGCCCACTTCGAGATAT 6904
OY 158 GIYAlaProProLeuGIYTYRGIYAlaProProLeuGIYTYRGIYThrProProLeuGIY 177
DB 6905 GAGGCCCACTTCGATATGAGCCCACTTCGATATGAGCCCACTTCGAG 6964
OY 178 TYRGIYAlaProProLeuGIYTYRGIYAlaProProAlaGIYTYRGIYAlaProProAla 197
DB 6965 TATGAGCCCACTTCGATATGAGCCCACTTCGAGATATGAGCCCACTTCG 7024
OY 198 GIYTYRAlaSerProAlaGIYSerGIYAlaArgProGIaUserThrAlaAlaGIa 217
DB 7025 GGATATGAGCCCACTTCGATATGAGCCCACTTCGATATGAGCCCACTTCGAG 7084
OY 218 AlaProGIaUser 222
DB 7085 GCTCCTGAAAACGAG 7099
```

RESULT 12

ID AAS70582 standard; cDNA; 894 BP.

AC AAS70582;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #6386.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSB-) HYSBQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG06395.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX Claim 1; SEQ ID NO 6386; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 894 BP; 204 A; 221 C; 213 G; 256 T; 0 other;

Alignment Scores:

Pred. No.:	6.95e-10	Length:	894
Score:	342.00 <td>Matches:</td> <td>69</td>	Matches:	69
Percent Similarity:	61.74% <td>Conservative:</td> <td>2</td>	Conservative:	2
Best Local Similarity:	60.00% <td>Mismatches:</td> <td>12</td>	Mismatches:	12
Query Match:	26.74% <td>Indels:</td> <td>32</td>	Indels:	32
DB:	23 <td>Gaps:</td> <td>2</td>	Gaps:	2

US-09-864-291-12 (1-235) x AAS70582 (1-894)

```
OY 130 GIUaGIYProProValaGIYTYRAlaSerProValaArgTYRGIYAlaProProLeuGIY 149
DB 168 GAGAGCCCGCTGAGATATGAGCCCACTTCGATATGAGCCCACTTCG 206
OY 150 TYRGIYAlaProProAlaGIYTYRGIYAlaProProLeuGIYTYRGIYAlaProProLeu 169
DB 207 -----TATGAGCCCACTTC 224
OY 170 GIYTYRGIYThrProProLeuGIYTYRGIYAlaProProLeuGIYTYRGIYAlaProPro 189
DB 225 GGATATGAGAACCCCACTTCGATATGAGCCCACTTCGATATGAGCCCACTTC 284
OY 190 AlaGIYanGIaGIYProProAlaGIYTYRAlaAlaSerProAlaGIYSerGIYAlaArg 209
DB 285 GCAGAAATAGAGCCCGCTGCGGATATGAGCCCACTTCGATATGAGCCAGG 344
OY 210 ProGIaGIaUser-----ThrAlaAlaGIa 218
DB 345 CCTGAGAAATCCAGGAGAAAGTGTGTCGCCAAGACTTGAGATGCCAGAACCATAGGG 404
OY 219 ProGIaanGIaAlaSerLeuProSerAlaSerSerGIaVal 233
DB 405 CCCCAAGAGGAGTCAAGCCCTGCTTGAGAGCTCCCAAGTC 449
```

RESULT 13

ID AAS70430 standard; cDNA; 1885 BP.

AC AAS70430;

DT 27-NOV-2002 (first entry)

XX Human bone remodelling gene #87.
 DR Bone remodelling; osteoporosis; human; gene; ss.
 XX Homo sapiens.
 OS
 XX US6426186-B1.
 PN
 XX 30-JUL-2002.
 PD
 XX 18-JAN-2000; 2000US-0484970.
 PF
 XX 18-JAN-2000; 2000US-0484970.
 PR
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Jones KA, Volkmuth W, Walker MG;
 PI
 XX WPI; 2002-673014/72.
 DR
 XX A combination of polynucleotides which are co-expressed with genes
 PT known to be involved in bone remodeling and osteoporosis are useful in
 PT an array for the diagnosis of bone remodeling and osteoporosis
 PT associated disorders -
 PT
 XX
 PS Claim 1; Column 247-250; 206pp; English.
 XS
 CC The invention relates to a combination comprising a number of
 CC substantially purified and isolated polynucleotides which are
 CC co-expressed with genes known to be involved in bone remodeling and
 CC osteoporosis. The invention is used to diagnose disorders associated
 CC with bone remodeling or osteoporosis. ABS70344-ABS70512 represent
 CC human bone remodelling genes of the invention.
 SX
 SQ Sequence 1885 BP; 384 A; 610 C; 476 G; 415 T; 0 other;
 Alignment Scores:
 Pred. No.: 1,91e-05 Length: 1885
 Score: 261.50 Matches: 76
 Percent Similarity: 41.70% Conservative: 22
 Best Local Similarity: 32.34% Mismatches: 74
 Query Match: 20.45% Indels: 63
 DB: 24 Gaps: 9
 US-09-864-291-12 (1-235) x ABS70430 (1-1885)
 QY 1 MetProPheAAspleuMetThraenLeuThrValGluGlnProValPheAlaAlaenphe 20
 DB 274 ATGCCATTATTATCTCATGAAGACTGTGAGATCAAGACGCCCTATTGTGTCAAACTAC 333
 QY 21 IleLyGlyThrTleGlnAlaAlaProTyrgLyTyrgLuglyGlnAlaThrPheLy 40
 DB 334 ATCAAGGGAACAGTGAAGCGGGAAGCGGAGTGGCTGGGGAAGGCTGCTCTTCAAG 393
 QY 41 leuVal.PheArgAsnGlyAspAlaIleGluPheAlaGluMetValLyAlaAlaSer 60
 DB 394 TTGACTTTTCAAGCGACGAGCGCGCCCATTTGATTCGACAGGAGTGTCTCGATCT 453
 QY 61 AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80
 DB 454 CAAGCTCCAGAGT----- 468
 QY 81 TyrValIleThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
 DB 469 -----GAAGTCCCGACGTGAGCCTATGAC 492
 QY 101 TyrGlyAlaProProAlaGlyTyrgLyAlaProProGlyTyrgLyAlaProProAla 120
 DB 493 TACTTTACATGCGCCAGCGG-----GCCATGTTCTATCCCG--- 531
 QY 121 GlyTyrgLyAlaGlnProValGlyAsnGluGlyProProValGlyTyrgAlaSerPro 140

DB 532 -----CCAGTGCAGAT-----GGAAATGACCCCTGCCCT 561
 QY 141 ValArgTyrgLyAlaProProLeuGlyTyrgLyAlaProProAlaGlyTyrgLyAlaPro 160
 DB 562 CCGGCTACCCCTATCCACCG-----CCCCACCTGAGTTCTATCCAGACCC 609
 QY 161 ProLeuGlyTyrgLyAlaProProLeuGlyTyrgLyThrProProLeuGlyTyrgLyAla 180
 DB 610 CCATGATGAGACGGGGCC-----ATGGGATACGTGACGCCCA----- 648
 QY 181 ProProLeuGlyTyrgLyAlaProProAlaGlyAsnGluGlyProProAlaGlyTyrg 200
 DB 649 CCACCG-----CCCTACCCCTGGCCCATGAGAACTCCGGTTCACGGGCC 693
 QY 201 AlaserProAlaGlySerGlyAlaArgProGlnGlnUserThrAlaAlaGlnAlaProGlu 220
 DB 694 CGATGTCCTCCACTCTGACGCGGAGCCAA-----GGCCGCAAGACGACGCC 744
 QY 221 AsnGlnAlaSerLeuProSerAlaSerSerSerGlnValHisSer 235
 DB 745 CAGCGCCTATTACAAACGAGGCAATCTCAAGAGTCTCATGTC 789
 RESULT 14
 AAF22363
 ID AAF22363 standard; cDNA; 1915 BP.
 XX
 AC AAF22363;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DR Human secreted protein gene 48 SEQ ID NO:58.
 XX
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; vulnary; gene therapy; neoplasm;
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischemia; angiogenesis; nervous system disorder; infection;
 KW Alzheimer's disease; ocular disorder; corneal infection; wound healing;
 KW skin aging; food additive; preservative; ss.
 OS Homo sapiens.
 XX
 PN WO200061748-A1.
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000WO-US08982.
 XX
 PR 09-APR-1999; 99US-0128696.
 PR 14-JAN-2000; 2000US-0176069.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX WPI; 2000-638566/61.
 DR P-PSDB; AAB63096.
 XX
 PT New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PT
 PS Claim 1; Page 429-430; 480pp; English.
 XX
 CC AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
 CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 CC cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;

fungicide; ophthalmological; and vulnerary. The polynucleotides and proteins can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. AAF22307 to AAF22315 and AAB6048 represent sequences used in the exemplification of the present invention.

XX Sequence 1915 BP, 401 A, 613 C, 481 G, 413 T, 7 other;

Alignment Scores:

Pred. No.:	2,94E-05	Length:	1915
Score:	258.00	Matches:	77
Percent Similarity:	42.13%	Conservative:	22
Best Local Similarity:	32.77%	Mismatches:	73
Query Match:	20.17%	Indels:	64
DB:	21	Gaps:	8

US-09-864-291-12 (1-235) x AAF22363 (1-1915)

```

OY 1 MetProPheApLeuMetThrAnuLeuThrValGluGlnProValPheAlaAlaSerPhe 20
DB 272 ATGCATTATTTATCTCATGAAGAAGCTGTGATCATCAAGCAGCCGATTTGGTCAAACTAC 331
OY 21 IleValGlyThrIleGlnAlaAlaProTyrglyGlyTTPGluGlyGlnAlaThrPheLys 40
DB 332 ATCAAGGAAACAGTAAAGCGGAGACCGGAGGTGGCTGGAAAGGCTCTGCTTCCATCAAG 391
OY 41 LeuValPheArgAsnGlyValAspAlaIleGluPheAlaGlnLeuMetValIleAlaSer 60
DB 332 TTGACTTTCACGCGGAGGCGGCCCATTTAGTTCCGACAGCGGATCTCCAGGTGCATCT 451
OY 61 AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTTPPheSerSerMetGlyIle 80
DB 452 CAAGCTCCAGAGGT----- 466
OY 81 TyrValIleThrGlyGluGlyValAsnMetCysThrProGlnMetProCysSerValIleVal 100
DB 467 -----GAAATGCCCAAGTGGAGCCCTATG 490
OY 101 TyrglyAlaProProAlaGlyTyrglyAlaProProProGlyTyrglyAlaProProAla 120
DB 491 TACTCTTACATGCCAGCGGSCYTATGTCATCCCGG----- 529
OY 121 GlyTyrglyAlaGlnProValGlyValGluGlyProProValGlyTyrglyAlaSerPro 140
DB 530 -----CGATGCCCAAT-----GGAATGATACCCCTGCT 559
OY 141 ValArgTyrglyAlaProProLeuGlyTyrglyAlaProProAlaGlyTyrglyAlaPro 160
DB 560 CCTGGCTACCCCTATCCACCG-----CCCAACCTGAGTTCTATCCAGAGCC 607
OY 161 ProLeuGlyTyrglyAlaProProLeuGlyTyrglyThrProProLeuGlyTyrglyAla 180
DB 608 CCCATGATGAGACGGGAGCC-----ATGGATATCGTGACAGCCCA----- 646
OY 181 ProProLeuGlyTyrglyAlaProProAlaGlyValGluGlyProProAlaGlyTyrgly 200
DB 647 CCACCG-----CCCTACCTGGGCGCCCATGAAACCTCCGGTACGCGGCGCC 691
OY 201 AlaSerProAlaGlySerGlyAlaArgProGlnGluSerThrAlaGlnAlaProGlu 220

```

```

DB 692 GATGCCCTCATCTCGACGCG-----GAGCGAAGCGCGAGAGAGCGC-CGC 741
OY 221 AsnGluAlaSerLeuProSerAlaSerSerSerGlnValIleSer 235
DB 742 CAGCGCCTATTACACCAAGCAGGCAATCTCTCAACAGCTTACATGCG 786

```

RESULT 15

ID AB273614 standard; cDNA, 1915 BP.

AC AB273614;

DT 12-MAY-2003 (first entry)

DE Secreted protein-encoding gene 334 cDNA clone HTLEMI6, SEQ ID NO:344.

XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antianaemic; vulnerary; gene; ss.

OS Homo sapiens.

PN WO200277013-A2.

PD 03-OCT-2002.

PF 26-MAR-2002; 2002WO-US09370.

PR 27-MAR-2001; 2001US-278650P.

PR 12-SEP-2001; 2001US-0950082.

PR 12-SEP-2001; 2001US-0950083.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR MPI; 2003-040578/03.

PT P-PSDB, ABR01280.

PT New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorder, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia -

PS Claim 21; Page 1332; 2474pp; English.

CC AB273281-AB273697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC AB273698-AB274687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein-encoding cDNA clone of the invention.

Sequence 1915 BP, 401 A, 613 C, 481 G, 413 T, 7 other;

Alignment Scores:

Pred. No.:	2,94e-05	Length:	1915
Score:	258.00	Matches:	77
Percent Similarity:	42.13%	Conservative:	22
Best Local Similarity:	32.77%	Mismatches:	73
Query Match:	20.17%	Indels:	64
DB:	25	Gaps:	8

US-09-864-291-12 (1-235) x AB273614 (1-1915)

```

QY      1 MetProPheAspLeuMetThrAsnLeuThrValGluGlnProValPheAlaAsnPhe 20
DB      272 ATGCCATTATCTCATGAAGACTGTGATCATCAAGACGCCCTATTGTGCAAACTAC 331
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      21 IleValGlyThrIleGlnAlaAlaProTyrGlyTyrGluGlyGlnAlaThrPheIle 40
DB      332 ATCAAGGGAACGATGAAGCGGAGCGGAGTGGCTGGAGAGGCTCTGCTTCTTCAAG 391
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      41 LeuValPheArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIlybAlaAsn 60
DB      392 TTACACTTCACGCGAGGGGGCGCCATTGAGTTGGACAGCGGATGCTCCAGTGGCATCT 451
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      61 AlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80
DB      452 CAAGCTCCAGAGGT-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      81 TyrValIleThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
DB      467 -----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      101 TyrGlyAlaProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaProProAla 120
DB      491 TACTCTTACATGCCCAAGGGGSCCTATGCTATCCCCG-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      121 GlyTyrGlyAlaGlnProValGlyAsnGluGlyProProValGlyTyrArgAlaSerPro 140
DB      530 -----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      141 ValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaPro 160
DB      560 CCTGGCTACCCCTATCCACCG-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      161 ProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAla 180
DB      608 CCCATATGACCGGGCC-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      181 ProProLeuGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAlaGlyTyrArg 200
DB      647 CCACCG-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      201 AlaSerProAlaGlySerGlyAlaArgProGlnGluSerThrAlaAlaGlnAlaProGlu 220
DB      692 GATGTCCCTCCACTCCTGCAAGCC-----
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      221 AsnGluAlaSerLeuProSerAlaSerSerSerGlnValHisSer 235
DB      742 CAGCGCTATTACCAACCCAGGCAATCTTCAACAAGTCTTACATGCC 786
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: December 16, 2003, 18:09:00
 Job time : 261.834 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 17:36:35 ; Search time 57.0347 Seconds
(without alignment)
1818.631 Million cell updates/sec

Title: US-09-864-291-12

Perfect score: 1279

Sequence: 1 MPDLMNTLVTEQVFAANF.....AQAPENBASLPSSSSQVHS 235

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 segs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlp
-O=/cgn2_1/USPRO_epool_p/US09864291/runat_15122003_160859_23332/app_query.fasta_1.846
-DB=Issued Patents NA -OFMT=fastap -SUFFIX=rml -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09864291 @cgn2_1_115 @runat_15122003_160859_23332 -NCPU=6 -ICPU=3
-NO_MMAP -LARGESQUERY -NEG_SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=10 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FAPOP=6
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	261.5	20.4	1885	4	US-09-484-970B-87
2	206.5	16.1	756	4	US-08-642-255-50
3	196	15.3	654	3	US-08-998-416-1144
4	192	15.0	4403765	3	US-09-103-840A-2
5	192	15.0	4411529	3	US-09-103-840A-1
6	188.5	14.7	432	1	US-08-642-255-48
7	186	14.5	2144	3	US-08-834-306-15
8	186	14.5	2144	3	US-08-993-674A-15
9	186	14.5	2144	4	US-09-256-976-15
10	182.5	14.3	744	4	US-09-183-861-60
11	182.5	14.3	744	4	US-09-022-756-60
12	182.5	14.3	744	4	US-09-551-974A-60

13	181	14.2	1235	2	US-08-557-309B-53	Sequence 53, Appl
14	166	13.0	2830	2	US-09-010-928B-1	Sequence 1, Appl
15	164.5	12.9	2824	2	US-09-010-928B-3	Sequence 3, Appl
16	162.5	12.7	2344	4	US-09-347-878-31	Sequence 31, Appl
17	162.5	12.7	3171	4	US-09-169-768-19	Sequence 19, Appl
18	162	12.7	810	1	US-08-642-255-60	Sequence 60, Appl
19	159.5	12.5	774	3	US-08-956-307B-12	Sequence 12, Appl
20	159.5	12.5	778	3	US-08-956-307B-11	Sequence 11, Appl
21	159.5	12.5	2428	1	US-08-445-050-1	Sequence 1, Appl
22	159.5	12.5	2428	1	US-08-204-691-1	Sequence 1, Appl
23	159.5	12.5	2428	4	US-09-355-295B-2	Sequence 2, Appl
24	159.5	12.5	2487	3	US-08-370-223-12	Sequence 12, Appl
25	159.5	12.5	2493	1	US-07-977-434-5	Sequence 5, Appl
26	159.5	12.5	2493	1	US-08-458-819-5	Sequence 5, Appl
27	159.5	12.5	2493	5	PCT-US91-07035-5	Sequence 5, Appl
28	159.5	12.5	3018	1	US-08-347-718B-3	Sequence 3, Appl
29	159.5	12.5	3018	1	US-08-482-262-3	Sequence 3, Appl
30	159.5	12.5	3018	6	5200183-1	Sequence 6, Appl
31	159.5	12.5	11531	1	US-08-068-945A-1	Sequence 1, Appl
32	159.5	12.5	11531	1	US-08-442-806-1	Sequence 1, Appl
33	159.5	12.5	11531	4	US-09-385-285B-1	Sequence 1, Appl
34	157.5	12.3	4167	4	US-09-169-768-12	Sequence 12, Appl
35	157	12.3	2681	3	US-08-928-213B-7	Sequence 7, Appl
36	156	12.2	3170	4	US-09-169-768-1	Sequence 1, Appl
37	156	12.2	3171	4	US-09-169-768-15	Sequence 15, Appl
38	156	12.2	3181	1	US-08-655-086-1	Sequence 1, Appl
39	156	12.2	3349	4	US-09-169-768-13	Sequence 13, Appl
40	156	12.2	3531	4	US-09-169-768-7	Sequence 7, Appl
41	156	12.2	3541	4	US-09-169-768-9	Sequence 9, Appl
42	155.5	12.2	2496	1	US-08-073-384C-2	Sequence 2, Appl
43	155.5	12.2	2496	1	US-08-254-359A-2	Sequence 2, Appl
44	155.5	12.2	2496	1	US-08-483-043-2	Sequence 2, Appl
45	155.5	12.2	2496	1	US-08-481-238-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-484-970B-87
; Sequence 87, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmutch, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 87
; LENGTH: 1885
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6426186 232968.10CB1
US-09-484-970B-87

Alignment Scores:

Pred. No.: 2.69e-12 Length: 1885
Score: 261.50 Matches: 76
Percent Similarity: 41.70% Conservative: 22
Best Local Similarity: 32.34% Mismatches: 74
Query Match: 20.45% Indels: 63
DB: 4 Gaps: 9

US-09-864-291-12 (1-235) x US-09-484-970B-87 (1-1885)

Oy 1 MePProheapleuMetThraenLeThVaGluGlnProValaAaasphe 20
Db 274 ATGCCATTATCTCATGAAGAACTGTGAAGATCAAGACGCCGATTGTGCTCAACTAC 333

APPLICANT: FRASER, Claire M.
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertam I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A5556-3/BIR
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic"
US-08-642-255-48

Alignment Scores:
Pred. No.: 3.04e-07 Length: 432
Score: 188.50 Matches: 70
Percent Similarity: 52.52% Conservative: 35
Best Local Similarity: 50.36% Mismatches: 35
Query Match: 14.74% Indels: 32
DB: 1 Gaps: 14

US-09-864-291-12 (1-235) x US-09-103-840A-1 (1-4411529)
QY 95 ProCyServal111eVal1YrG1yAlaProPro1ag1YrG1yAlaProPro----- 112
DB 4130468 CCGGAGCATGCTG-----CCGCTGGCTACCCGGTTGAACACACCGCC 4130512
QY 113 -----ProG1YrG1yAlaProPro1ag1YrG1yAlaGlnProVal1G1yAnG1u 130
DB 4130513 GTGGCGCGGGATATGCG-----CCGGCGGGATATCGCGCTACCC----- 4130554
QY 131 G1yAlaProPro1ag1YrG1yAlaProPro1ag1YrG1yAlaProPro1ag1Yr 150
DB 4130555 GGTACACACACCGGGTAC-----GGCCCGCGGGTATGTCGCCCGCC---AGCTAT 4130605
QY 151 G1yAlaProPro1ag1YrG1yAlaProPro1ag1YrG1yAlaProPro1ag1Yr 170
DB 4130606 GGC-----CCCGCGCTGGCTATAGT---CCACCCCTCGGCTACCCCGCGCACCGCGGC 4130659
QY 171 TyrG1YrProPro1ag1YrG1yAlaProPro1ag1YrG1yAlaProPro1a 190
DB 4130660 TCGGCGCCACCGCC---GCTATAGC---CCACCGCTCGGCTATAGCGCCACCGGTCCGC 4130713
QY 191 G1yAnG1uG1yProPro----- 196
DB 4130714 CCGGCGCGGTCAACACGAGATATATCCGCTGCCGCGCTTGAAGCATATCTTC 4130773
QY 197 -----AlaG1Yr---Arg1asePro 203
DB 4130774 AACGGCGCGGTGCGGTACATCGCGCTAACCG 4130806

RESULT 6
US-08-642-255-48/c
Sequence 48, Application US/08642255
Patent No. 5773249
GENERAL INFORMATION:
APPLICANT: CAPPELLO, Joseph
TITLE OF INVENTION: High Molecular Weight Collagen-Like
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSES: FLHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertam I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A5556-3/BIR
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic"
US-08-642-255-48

Alignment Scores:
Pred. No.: 3.04e-07 Length: 432
Score: 188.50 Matches: 70
Percent Similarity: 52.52% Conservative: 35
Best Local Similarity: 50.36% Mismatches: 35
Query Match: 14.74% Indels: 32
DB: 1 Gaps: 14

US-09-864-291-12 (1-235) x US-08-642-255-48 (1-432)
QY 82 Val11eThrG1yG1yAnMeTcYrTh-ProG1nMeTcProCyServal11eVal1Yr 101
DB 400 GTCTGCTGGGCTATGACCTGCGCGG---CTTCCAGGTCC----- 360
QY 102 G1yAlaPro-----Pro1ag1YrG1yAlaPro-----ProProG1YrG1yAla 117
DB 359 GAGACACCGGTGACGACCGGTGACGAGGCACTGCGCGGCTGAGAGTCCCGAGACA 300
QY 118 Pro-----Pro1ag1YrG1yAlaGlnProVal1G1yAnG1uG1yProPro1a 135
DB 299 CCGGTGACGACCGGTGACGAGGCA---CTTGGCGGCTGACGAGTCC---GGA 249
QY 136 TyrArg1aseProValArg1YrG1yAlaPro-----Pro1ag1YrG1yAlaPro 153
DB 248 GACCCGGTGAACACCGCGGTGACGAGGCACTGCGCGGCTGAGAGTCCCGAGACCC 189
QY 154 -----Pro1ag1YrG1yAlaPro-----Pro1ag1YrG1yAlaPro----- 167
DB 188 GGTGACACGCGGTGACGAGGCACTGCGCGGCTGAGAGTCCCGAGACCCCGGTGAGA 129
QY 168 Pro1ag1YrG1YrPro-----Pro1ag1YrG1yAlaPro-----Pro1eu 183
DB 128 CCAGCGGTGACGAGGCACTGCGCGGCTGACGAGTCCCGAGACCCCGGTGAGACGCC 69
QY 184 G1YrG1yAlaPro-----Pro1ag1YrG1yAnG1uG1yPro-----Pro1ag1Yr 198
DB 68 GGTCCAGCGGACCTTTCGGGCTGCTGGGCACTGCGGCTTTCGAGACTGCTGGG 12

RESULT 7
US-08-834-306-15
Sequence 15, Application US/08834306
Patent No. 6054135
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: Ske1y, Yasir A.W.
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSES: Houghton, Raymond L.
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2144 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-834-306-15

Alignment Scores:
Pred. No.: 3,54e-06 Length: 2144
Score: 186.00 Matches: 51
Percent Similarity: 40.91% Conservative: 3
Best Local Similarity: 38.64% Mismatches: 71
Query Match: 14.54% Indels: 7
DB: 3 Gaps: 1

US-09-864-291-12 (1-235) x US-08-834-306-15 (1-2144)

QY 84 ThrGlyGluGlyAanMetCyethrProGlnMetProCyseVal----- 98
DB 1211 ACCCGCGAAGGCTGCTGCTGCAACCGCGAAGCGCGCTGCCACCGCGAAGACCGCTGC 1270
QY 99 ---11eValTyrgLy-AlaProProAlaGlyTyrgLyAlaProProGlyTyrgLyAl 117
DB 1271 TGCACCGCGAAGGCTGCTGCTGCAACCGCGAAGCGCTGCTGCCACCGCGAAGCGCTGC 1330
QY 117 aProProAlaGlyTyrgLyAlaGlnProValGlyAanGluGlyProProValGlyTyra 137
DB 1331 TCCACCGCGAAGACCGCTGCTGCCACCGCGAAGACCGCTGCTGCCACCGCGAAGCGCTGC 1390
QY 137 gAlaserProValaGlyTyrgLyAlaProProLeuGlyTyrgLyAlaProProAlaGlyTy 157
DB 1391 TGCTCAACCGCGAAGGCGCGCTGCTGCCACCGCGAAGCGCGCTGCTGCCACCGCGAAGGC 1450
QY 157 rGlyAlaProProLeuGlyTyrgLyAlaProProLeuGlyTyrgLyThrProProLeuG 177
DB 1451 CGCTCTGCAACCGCGAAGCGCGCTGCTGCCACCGCGAAGCGCTGCTGCCACCGCGAA 1510
QY 177 YTyrgLyAlaProProLeuGlyTyrgLyAlaProProAlaGlyAanGluGlyProProAl 197
DB 1511 GGCGCGCTGCTGCCACCGCGAAGGCGCTGCTGCCACCGCGAAGGCGCTGCTGCCACCGCGC 1570
QY 197 aGlyTyraGAlaserProAlaGlySerGlyAla 208
DB 1571 GAAGGCTGCTGCTGCCCGTTCGAAAGAGGCT 1604

RESULT 8
US-08-993-674A-15
; Sequence 15, Application US/08993674A

Patent No. 6228372
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2144 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-993-674A-15

Alignment Scores:
Pred. No.: 3,54e-06 Length: 2144
Score: 186.00 Matches: 51
Percent Similarity: 40.91% Conservative: 3
Best Local Similarity: 38.64% Mismatches: 71
Query Match: 14.54% Indels: 7
DB: 3 Gaps: 1

US-09-864-291-12 (1-235) x US-08-993-674A-15 (1-2144)

QY 84 ThrGlyGluGlyAanMetCyethrProGlnMetProCyseVal----- 98
DB 1211 ACCCGCGAAGGCTGCTGCTGCAACCGCGAAGCGCGCTGCCACCGCGAAGACCGCTGC 1270
QY 99 ---11eValTyrgLy-AlaProProAlaGlyTyrgLyAlaProProGlyTyrgLyAl 117
DB 1271 TGCACCGCGAAGGCTGCTGCTGCAACCGCGAAGCGCTGCTGCCACCGCGAAGCGCTGC 1330
QY 117 aProProAlaGlyTyrgLyAlaGlnProValGlyAanGluGlyProProValGlyTyra 137
DB 1331 TCCACCGCGAAGACCGCTGCTGCCACCGCGAAGACCGCTGCTGCCACCGCGAAGCGCTGC 1390
QY 137 gAlaserProValaGlyTyrgLyAlaProProLeuGlyTyrgLyAlaProProAlaGlyTy 157
DB 1391 TGCTCAACCGCGAAGGCGCGCTGCTGCCACCGCGAAGGCGCGCTGCTGCCACCGCGAAGGC 1450
QY 157 rGlyAlaProProLeuGlyTyrgLyAlaProProLeuGlyTyrgLyThrProProLeuG 177
DB 1451 CGCTCTGCAACCGCGAAGCGCGCTGCTGCCACCGCGAAGCGCTGCTGCCACCGCGAA 1510
QY 177 YTyrgLyAlaProProLeuGlyTyrgLyAlaProProAlaGlyAanGluGlyProProAl 197

Db 1511 GGCGCTGCTCCACCGCGAAGGCTGCTGCTCCACCGCGAAGGCTGCTCCACCGCG 1570
Qy 197 AGLYTARGAlaserProAlaglysergYAla 208
Db 1571 GAAGGCTGCTGCTCCCGCTTGGAAGAGGCT 1604

RESULT 9

US-09-256-976-15
Sequence 15, Application US/09256976
Patent No. 641933
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
OF T. CRUZI INFECTION
FILE REFERENCE: 210121.422C3
CURRENT APPLICATION NUMBER: US/09/256,976
CURRENT FILING DATE: 1999-02-24
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 2144
TYPE: DNA
ORGANISM: Trypanosoma cruzi
US-09-256-976-15

Alignment Scores:

Pred. No.:	3,54e-06	Length:	2144
Score:	186.00	Matches:	51
Percent Similarity:	40.91%	Conservative:	3
Best Local Similarity:	38.64%	Mismatches:	71
Query Match:	14.54%	Indels:	7
DB:	4	Gaps:	1

US-09-864-291-12 (1-235) x US-09-256-976-15 (1-2144)

Qy 84 ThrtgYgluYlAgnMetCysThrProGlnMeCProCysSerVal----- 98
Db 1211 ACCCGGAGAGGCTGCTGCTGACCCGCGAAGCGCGTCTCCACCGCGAAGCGCTGC 1270
Qy 99 ---11eValYrgly-A1aProProAlaglyTyrGlyAlaProProProGlyTyrGlyAl 117
Db 1271 TGCACCGCGAAGGCTGCTGCACTGCCAAGGCTGCTGCCACCGCGAAGGCTGCTGC 1330
Qy 117 AProProAlaglyTyrGlyAlaGlnProValGlyAgnGluGlyProProValGlyTyrAr 137
Db 1331 TCCACCGCGAAGCGCTGCTGCCACCGCGAAGCGCTGCTCCACCGCGAAGGCTGC 1390
Qy 137 gAlaSerProValArGlyTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaglyTyr 157
Db 1391 TGTCTCACCGCGAAGCGCTGCTGCCACCGCGAAGCGCTGCTCCACCGCGAAGGC 1450
Qy 157 rGlyAlaProProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuG 177
Db 1451 CGGTGCTGACCGCGAAGCGCGTCTGCCACCGCGAAGGCTGCTGCCACCGCGAAG 1510
Qy 177 YTYrGlyAlaProProLeuGlyTyrGlyAlaProProAlaglyAgnGluGlyProProAl 197
Db 1511 GCGCGTGTCTCCACCGCGAAGGCTGCTGCCACCGCGAAGGCTGCTGCCACCGCG 1570
Qy 197 aGlyTyrArGAlaSerProAlaglysergYAla 208
Db 1571 GAAGGCTGCTGCTGCTCCCGCTTGGAAGAGGCT 1604

RESULT 10

US-09-183-861-60/c
Sequence 60, Application US/09183861
Patent No. 6365165
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skelky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/183,861
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/022,765
FILING DATE: 12-FEB-1998
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 744 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-183-861-60

Alignment Scores:

Pred. No.:	1.81e-06	Length:	744
Score:	182.50 <td>Matches:</td> <td>63</td>	Matches:	63
Percent Similarity:	39.18%	Conservative:	4
Best Local Similarity:	36.84%	Mismatches:	31
Query Match:	14.27%	Indels:	74
DB:	4	Gaps:	11

US-09-864-291-12 (1-235) x US-09-183-861-60 (1-744)

Qy 100 ValTYrGlyAlaProProAlagly-----TYrGlyAlaProPro 112
Db 583 GTTACGGCGCGTCTCTTAGGGCCATCGCGCATTTTCTGTGTATGCGCTCTCT 524
Qy 113 ProGly-----TYrGlyAlaProProAlagly----- 121
Db 523 TAGGGCCATCGCGCATTTTCTGTGTATGCGCGTCTCTTAGGGCCATCGCGCAT 464
Qy 122 -----TYrGlyAlaGlnProValGlyAgnGluGlyProProValGlyTyrArGAla 138
Db 463 TTTTCTGTGTATGCGCGT-----CTCTCTTAGGGCCATCGCGCAT 425
Qy 139 SerProValArGlyTyrGlyAlaProProLeuGly-----TYrGly 151
Db 424 CATTTTCTGTGTATGCGCGTCTCTCTTAGGGCCATCGCGCATTTTCTGTGTATGCG 365
Qy 152 AlAProProAlagly-----TYrGlyAlaProProLeuGly----- 163
Db 364 CGTCTCTTAGGGCCATCGCGCATTTTCTGTGTATGCGCGTCTCTCTTAGGGCCAT 305
Qy 164 -----TYrGlyAlaProProLeuGly----- 170

DB 304 CGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAGGGCCATGCGGTATTTTCTGTG 245
QY 171 TTTGTATTTTCTGTGTAGCGCGTCTCTCTTAGGGCCATGCGGTATTTTCTGTG 183
DB 244 TAGGCGCGTCTCTCTTAGGGCCATGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAG 185
QY 164 GY-----TyrGlyAlaProProAlaGlyYasnGlu----- 193
DB 164 GCGCATGCGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAGGGCCATGCGGTATTTTCTGTG 126
QY 194 GlyProProAlaGlyTyrArgAlaSerProAla 204
DB 125 GCACCGCGCGGTCTCTTAGGGCCATGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAG 93

RESULT 11
US-09-022-765-60/c
Sequence 60, Application US/09022765
Patent No. 6375955
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE THERAPY AND
NUMBER OF SEQUENCES: 87
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,765
FILING DATE: 12-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.420C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6011
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 744 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-022-765-60

Alignment Scores:
Pred. No.: 1 81e-06 Length: 744
Score: 182.50 Matches: 63
Percent Similarity: 39.18% Conservative: 4
Best Local Similarity: 36.84% Mismatches: 31
Query Match: 14.27% Indels: 74
DB: 4 Gaps: 11

US-09-864-291-12 (1-235) x US-09-022-765-60 (1-744)

QY 100 ValTyrGlyAlaProProAlaGly-----TyrGlyAlaProPro 112
DB 583 GTGTAGCGCGGTCTCTCTTAGGGCCATGCGGTATTTTCTGTGTAGCGCGTCTCTCT 524
QY 113 ProGly-----TyrGlyAlaProProAlaGly----- 121

DB 523 TAGGCGCATGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAGGGCCATGCGGTAT 464
QY 122 -----TyrGlyAlaGlnProValGlyYasnGluGlyProProValGlyTyrArgAla 138
DB 463 TTTTCTGTGTAGCGCGGT-----CTCTCTTAGGGCCATGCGGTATTTTCTGTG 425
QY 139 SerProValAlaGlyTyrGlyAlaProProAlaGly-----TyrGly 151
DB 424 CATTTTCTGTGTAGCGCGGTCTCTCTTAGGGCCATGCGGTATTTTCTGTGTAGCGG 365
QY 152 AlaProProAlaGly-----TyrGlyAlaProProAlaGly--- 163
DB 364 GTCTCTCTTAGGGCCATGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAGGGCCAT 305
QY 164 -----TyrGlyAlaProProAlaGly----- 170
DB 304 CGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAGGGCCATGCGGTATTTTCTGTG 245
QY 171 TTTGTATTTTCTGTGTAGCGCGTCTCTCTTAGGGCCATGCGGTATTTTCTGTG 183
DB 244 TAGGCGCGTCTCTCTTAGGGCCATGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAG 185
QY 184 GY-----TyrGlyAlaProProAlaGlyYasnGlu----- 193
DB 184 GCGCATGCGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAGGGCCATGCGGTATTTTCTGTG 126
QY 194 GlyProProAlaGlyTyrArgAlaSerProAla 204
DB 125 GCACCGCGCGGTCTCTTAGGGCCATGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAG 93

RESULT 12
US-09-551-974A-60/c
Sequence 60, Application US/09551974A
Patent No. 6500437
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Webb, John R.
APPLICANT: Dillon, Davin C.
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
FILE REFERENCE: 210121.420C5
CURRENT APPLICATION NUMBER: US/09/551,974A
CURRENT FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 101
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 60
LENGTH: 744
TYPE: DNA
ORGANISM: Leishmania chagasi
US-09-551-974A-60

Alignment Scores:
Pred. No.: 1 81e-06 Length: 744
Score: 182.50 Matches: 63
Percent Similarity: 39.18% Conservative: 4
Best Local Similarity: 36.84% Mismatches: 31
Query Match: 14.27% Indels: 74
DB: 4 Gaps: 11

US-09-864-291-12 (1-235) x US-09-551-974A-60 (1-744)

QY 100 ValTyrGlyAlaProProAlaGly-----TyrGlyAlaProPro 112
DB 583 GTGTAGCGCGGTCTCTCTTAGGGCCATGCGGTATTTTCTGTGTAGCGCGTCTCTCT 524
QY 113 ProGly-----TyrGlyAlaProProAlaGly----- 121
DB 523 TAGGCGCATGCGGTATTTTCTGTGTAGCGCGTCTCTCTTAGGGCCATGCGGTAT 464
QY 122 -----TyrGlyAlaGlnProValGlyYasnGluGlyProProValGlyTyrArgAla 138

Db 463 TTTTCGTGTACGCGCT-----CCTCTTAGGCGCATCGCGT 425
QY 139 SerProValArgTyrGlyAlaProProLeuGly-----TyrGly 151
Db 424 CATTTTCTGTGTACGCGCTCTCTTAGGCGCATCGCGTATTTTCTGTGTACGCG 365
QY 152 AlaProProAlaGly-----TyrGlyAlaProProLeuGly--- 163
Db 364 CGTCTCTCTTAGGCGCATCGCGCATTTTCTGTGTATGCGCGTCTCTTAGGCGCAT 305
QY 164 -----TyrGlyAlaProProLeuGly----- 170
Db 304 CGCGTATTTTCTGTGTACGCGCTCTCTTAGGCGCATCGCGTATTTTCTGTG 245
QY 171 TyrGlyThrProProLeuGly-----TyrGlyAlaProProLeu 183
Db 244 TAGGCGCTCTCTTAGGCGCATCGCGTCTCTGTGTGTACGCGCTCTCTTAG 185
QY 184 Gly-----TyrGlyAlaProProAlaGlyAenglyu----- 193
Db 184 GCGCATCGCGTCTATTTTCTGTGTATGCGCATCTCTCTCGGCGCA-GAGTGTTCATC 126
QY 194 GlyProProAlaGlyTyrArgAlaSerProAla 204
Db 125 GCAACGCGCGCTGTCTCGGCGCGCACACGCGCG 93

RESULT 13

US-08-557-309B-53
Sequence 53, Application US/08557309B
Patent No. 5916572

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yaeli A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121,422
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 53:

SEQUENCE CHARACTERISTICS:
LENGTH: 1235 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-557-309B-53

Alignment Scores:

Pred. No.: 4.48e-06 Length: 1235
Score: 181.00 Matches: 51
Percent Similarity: 41.04% Conservative: 4
Best Local Similarity: 38.06% Mismatches: 70

Query Match: 14.15% Indels: 9
DB: 2 Gaps: 2
US-09-864-291-12 (1-235) x US-08-557-309B-53 (1-1235)

QY 84 ThrGlyGlyGlyAlaMetCysThrProGlnMetProCysSerVal-IleValTyrGlyAla 103
Db 301 ACCCGGAAAGGCTGTGTGTACACCGGAAAGCCGCTGTCTCAACCGCGAAGAGCTTC 360
QY 103 AProPro-----AlaGlyTyrGlyAlaProProProGlyTyrGlyAla 117
Db 361 TGCACCGGAAAGGCTGTGTGTACACCGGAAAGCTGTCTCAACCGCGAAGAGCTTC 420
QY 117 AProProAlaGlyTyrGlyAlaGlnProValGlyAenglyuGlyProProValGlyTyr 137
Db 421 TCACCGCGGAAAGAGCCGCTGTCTCAACCGCGAAGCCGCTGTCTCAACCGCGAAGCTTC 480
QY 137 GalSerProValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyr 157
Db 481 TGTCTCAACCGGAAAGGCGCTGTCTCAACCGCGAAGCCGCTGTCTCAACCGCGAAGCTTC 540
QY 157 rGlyAlaProProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuG 177
Db 541 CGCTGTGTGACCGCGAAGCGCGCTGTGTGTACACCGCGAAGCTGTCTCAACCGCGAA 600
QY 177 rTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyAenglyuGlyProProAl 197
Db 601 GCGCGCTGTCTCAACCGCGAAGCGCTGTCTCAACCGCGAAGCTGTCTCAACCGCG 660
QY 197 rGlyTyrArgAlaSerProAlaGlySerGlyAlaArgPro 210
Db 661 G-----GAGGCTGTCTGTCTCGCTCGTGTGAAAGAGCTT 694

RESULT 14

US-09-010-928B-1
Sequence 1, Application US/09010928B
Patent No. 5994099

GENERAL INFORMATION:

APPLICANT: Lewis, Randolph V
APPLICANT: Hayashi, Cheryl Y
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 GATHEHOUSE RD, SUITE 500E
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M

REGISTRATION NUMBER: 28977
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2830 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: -

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 17:58:55 ; Search time 229.854 Seconds

(without alignment) 3398.008 Million cell updates/sec

Title: US-09-864-291-12

Perfect score: 1279

Sequence: 1 MPDMLNTLTVGEVPANF.....AQAPENASLPSASSQVHS 235

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 440344

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODL=frame+ p2n.model -DEV=xip
-Q=/cgn2_1/USPRO_pool_P/US09864291/rnatc_15122003.160900.23360/app_query.fasta_1.846
-DB=Published Applications NA -OPMT=fastac -SUPFIX=rmpb -MINMATCH=0.1
-LOOCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=D10sum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09864291 @CCN 1.1.108 @rnatc_15122003.160900.23360
-NCPU=6 -ICPU=3 -NO MAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DISPLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

No.	Score	Query Match	Length	ID	Description

1	826	64.6	467	11	US-09-764-891-1230	Sequence 1230, Ap
2	761	59.5	436	9	US-09-864-761-11561	Sequence 31561, A
3	705	55.1	7099	11	US-09-764-891-7570	Sequence 7570, Ap
4	256	20.0	511	13	US-10-029-386-24981	Sequence 24981, A
5	245	19.2	471	9	US-09-864-761-15029	Sequence 15029, A
6	226	17.7	511	13	US-10-029-386-24981	Sequence 24981, A
7	211	16.5	541	13	US-10-029-386-11265	Sequence 11265, A
8	202.5	15.8	1743	10	US-09-887-576-788	Sequence 788, Ap
9	197.5	15.4	2430	15	US-10-156-761-2531	Sequence 2531, Ap
10	197.5	15.4	9025608	15	US-10-156-761-1	Sequence 1, Appl
11	192.5	15.0	866	13	US-10-029-386-22839	Sequence 22839, A
12	192	15.0	321	11	US-09-764-891-1356	Sequence 1356, A
13	191	14.9	432	11	US-09-818-995-33536	Sequence 33536, A
14	185	14.5	780	13	US-10-029-386-20220	Sequence 20220, A
15	185	14.5	2016	10	US-09-938-842A-2004	Sequence 2004, Ap
16	184	14.4	507	11	US-09-770-861-263	Sequence 263, Ap
17	182.5	14.3	744	9	US-09-874-923-60	Sequence 60, Appl
18	182.5	14.3	744	10	US-09-991-496-60	Sequence 60, Appl
19	182.5	14.3	3760	10	US-09-887-576-458	Sequence 458, Appl
20	178.5	14.0	534	10	US-09-938-842A-968	Sequence 968, Ap
21	178.5	14.0	731	9	US-09-770-149-103	Sequence 103, Appl
22	178.5	14.0	1909	13	US-10-094-749-1166	Sequence 1166, Ap
23	176	13.8	959	13	US-10-029-386-22751	Sequence 22751, A
24	176	13.8	9025608	15	US-10-156-761-1	Sequence 1, Appl
25	175	13.7	1743	15	US-10-128-714-7526	Sequence 7526, Ap
26	173	13.5	2164	15	US-10-128-714-1526	Sequence 1526, Ap
27	173	13.5	2217	15	US-10-128-714-6526	Sequence 6526, Ap
28	173	13.5	4164	15	US-10-128-714-5526	Sequence 5526, Ap
29	173	13.5	4217	15	US-10-128-714-5526	Sequence 5526, Ap
30	172	13.4	1967	9	US-09-864-761-5097	Sequence 5097, Appl
31	170.5	13.3	2485	15	US-10-153-668-91	Sequence 91, Appl
32	169.5	13.3	685	13	US-10-029-386-20888	Sequence 20888, A
33	169	13.2	593	13	US-10-027-632-290832	Sequence 290832, A
34	169	13.2	593	14	US-10-027-632-290832	Sequence 290832, A
35	169	13.2	1674	15	US-10-156-761-3052	Sequence 3052, Ap
36	168.5	13.2	1040	13	US-10-029-386-26033	Sequence 26033, A
37	168	13.1	541	13	US-10-029-386-11265	Sequence 11265, A
38	166	13.0	631	13	US-10-029-386-20805	Sequence 20805, A
39	166	13.0	2108	10	US-09-962-832-225	Sequence 225, Appl
40	166	13.0	2709	15	US-10-156-761-3543	Sequence 3543, Ap
41	165.5	12.9	335	11	US-09-913-514-26	Sequence 26, Appl
42	165.5	12.9	648	9	US-09-864-761-21820	Sequence 21820, A
43	165.5	12.9	648	13	US-10-029-386-20230	Sequence 20230, A
44	165	12.9	409	10	US-09-960-352-6984	Sequence 6984, Ap
45	163	12.7	1470	14	US-10-029-180-75	Sequence 75, Appl

ALIGNMENTS

RESULT 1
US-09-764-891-1230
; Sequence 1230, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1230
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: (421)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1230
Alignment Scores:

Pred. No.: 2.28e-71 Length: 467
 Score: 826.00 Matches: 149
 Percent Similarity: 96.77% Conservative: 1
 Best Local Similarity: 96.13% Mismatches: 4
 Query Match: 64.58% Indels: 1
 DB: 11 Gaps: 0

US-09-864-291-12 (1-235) x US-09-764-891-1230 (1-467)

Qy 44 ArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIleValAlaIleValAla 63
 Db 3 AGAATGTAGAGGTGCTGATGTAATGTCACAGTGTGTGTAAGCTGCTGCTGCTGCC 62
 Qy 64 ArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIleValIle 83
 Db 63 CGAGATTTTCACCTTAAACCTTAAATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
 Qy 84 ThrGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleValIleValIle 103
 Db 123 ACTGGGAGAGGATATGTGACCTCCACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 182
 Qy 104 ProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaProProAlaGlyTyrGly 123
 Db 183 CCACTGAGAGATATGAGCCCACTCCGATACGAGCCCACTGAGATATGAG 242
 Qy 124 AlaGlnProValGlyAsnGluGlyProProValGlyTyrArgAlaSerProValArgTyr 143
 Db 243 GCCCAACCCGTAGGAATGAGAGCCCGCTGTGTGAGACAGAGCTCACTGTGCTGATAT 302
 Qy 144 GlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProLeuGly 163
 Db 303 GAGCCCCCACTCTTGATGATGAGAGCCCACTGAGATATGAGAGCCCACTCTTGAG 362
 Qy 164 TyrGlyAlaPro-ProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAlaProProLe 183
 Db 363 TAGGAGCCCACTCTTGATGATGAGAGCCCACTCTGATATGAGAGCCCACTTNT 422
 Qy 183 uGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAla 197
 Db 423 CGGATATGAGAGCCCACTGAGGAATGAGAGCCCGCTTGC 465

RESULT 2

US-09-864-761-31561
 ; Sequence 31561 Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wenheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecm1ca-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263,6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 31561
 ; LENGTH: 436
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO Z99716.4
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
 ; OTHER INFORMATION: EST HUMAN HIT: AW169980.1, EVALU8 8.00e-11
 ; OTHER INFORMATION: SWISSPROT HIT: P18616, EVALU8 3.00e-13
 ; US-09-864-761-31561

Alignment Scores:

Pred. No.: 4.09e-65 Length: 436
 Score: 761.00 Matches: 137
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 59.50% Indels: 0
 DB: 9 Gaps: 0

US-09-864-291-12 (1-235) x US-09-864-761-31561 (1-436)

Qy 99 IleValTyrGlyAlaProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAlaPro 118
 Db 3 ATGTCTATGAGAGCCCACTGAGATATGAGAGCCCACTCCGATATGAGAGCCCA 62
 Qy 119 ProAlaGlyTyrGlyAlaGlnProValGlyAsnGluGlyProProValGlyTyrArgAla 138
 Db 63 CTGCGAGATATGAGAGCCCACTGAGGAATGAGAGCCCGCTGTGTGATACAGAGCC 122
 Qy 139 SerProValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGly 158
 Db 123 TCACCTGTGATATGAGAGCCCACTCTTGATATGAGAGCCCACTGAGATATGAG 182
 Qy 159 AlaProProLeuGlyTyrGlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyr 178
 Db 183 GCCCACTCTAGATATGAGAGCCCACTCTGATATGAGAGCCCACTCTCGATAT 242
 Qy 179 GlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyAsnGluGlyProProAlaGly 198
 Db 243 GAGAGCCCACTCTCGATATGAGAGCCCACTGAGGAATGAGAGCCCGCTGTGTGAG 302
 Qy 199 TyrArgAlaSerProAlaGlySerGlyAlaArgProGlnGlySerThrAlaAlaGlnAla 218
 Db 303 TACAGAGCTTCACTCTGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGCT 362
 Qy 219 ProGlyAsnGlyAlaSerLeuProSerAlaSerSerSerGlnValIleSer 235
 Db 363 CTGGAATGAGAGCTTCTCTTCCCTGCTGCTCTTCTTCAAGTCACTTCT 413

RESULT 3

```
US-09-764-891-7570
; Sequence 7570, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7570
; LENGTH: 7099
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-7570

Alignment Scores:
Pred. No.: 2.03e-58 Length: 7099
Score: 705.00 Matches: 125
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 55.12% Indels: 0
DB: 11 Gaps: 0

US-09-864-291-12 (1-235) x US-09-764-891-7570 (1-7099)

QY 98 ValIleValIYrGlyAlaProProAlaGlyYrGlyAlaProProGlyYrGlyAla 117
Db 6725 GTTATTGTCTATTGAGAGCCCACTGTCAGATATGAGCCCACTCCGGAATAGAGGCC 6784

QY 118 ProProAlaGlyYrGlyAlaGlnProValGlyYrGlyAlaGlnGlyYrGlyAla 137
Db 6785 CCACCTGAGAGATATGAGCCCACTGAGAAATGAGGCCCTGTCGATACAGA 6844

QY 138 AlaSerProValArgYrGlyAlaProProLeuGlyYrGlyAlaProProAlaGlyYr 157
Db 6845 GCTTCACCTGTCGATATGAGCCCACTGTCGATATGAGGCCCACTGTCGAGATAT 6904

QY 158 GlyAlaProProLeuGlyYrGlyAlaProProLeuGlyYrGlyAlaProProLeuGly 177
Db 6905 GAGAGCCCACTGTCGATATGAGCCCACTGTCGATATGAGCCCACTGTCGAG 6964

QY 178 TyrGlyAlaProProLeuGlyYrGlyAlaProProAlaGlyYrGlyAlaProProAla 197
Db 6965 TATGAGAGCCCACTGTCGATATGAGCCCACTGTCGAGAAATGAGGCCCTGTCG 7024

QY 198 GlyTyrAlaGlnAlaSerProAlaGlySerGlyAlaArgProGlnGlnSerThrAlaAla 217
Db 7025 GGAATACAGAGCTCCTGTCGATATGAGAGCCCACTGTCGAGATATGAGAGCCCA 7084

QY 218 AlaProGluAlaGlu 222
Db 7085 GCTCTGAAAACGAG 7099

RESULT 4
US-10-029-386-24981
; Sequence 24981, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: A60MCA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 24981
; LENGTH: 511
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; OTHER INFORMATION: MAP TO 297205.1
US-10-029-386-24981

Alignment Scores:
Pred. No.: 3.28e-16 Length: 511
Score: 256.00 Matches: 58
Percent Similarity: 47.14% Conservative: 8
Best Local Similarity: 41.43% Mismatches: 68
Query Match: 20.02% Indels: 6
DB: 13 Gaps: 2

US-09-864-291-12 (1-235) x US-10-029-386-24981 (1-511)

QY 102 GlyAlaProProAlaGlyYrGlyAlaProProGlyYrGlyAlaProProAlaGlyYr 119
Db 90 GGTCTCCCTCTCTCAGCAGAGCCCTCTCTCTCAGCAGAGTCCCTCTCTCAGCAGT 149

QY 120 AlaGlyYrGlyAlaGlnProValGlyYrGlyAlaGlnGlyYrGlyAlaGlnGlyYr 139
Db 150 GAGAGCTCCCTCTCTCAGCAGAGCTCCCTCTCTCAGCAGAGTCCCTCTCTCAGCAG 209

QY 140 ProValArgYrGlyAlaProProLeuGlyYrGlyAlaProProAlaGlyYrGlyAla 159
Db 210 CCTGAGAGCTCCCTCTCTCAGCAGAGTCCCTCTCTCAGCAGAGTCCCTCTCTCAGCAG 269

QY 160 ProProLeuGlyYrGlyAlaProProLeuGlyYrGlyAlaProProLeuGlyYrGly 179
Db 270 CCACCTGAGAGCTCTCTCTCTCAGCAGAGCTCTCTCTCTCAGCAGAGTCTCTCTCT 329

QY 180 AlaProProLeuGlyYrGlyAlaProProAlaGlyYrGlyAlaGlnGlyYrGlyAla 199
Db 330 CTTCTCAGCAGAGCTCTCTCTCTCAGCAGAGTCTCTCTCTCTCAGCAGAGTCTCTCT 389

QY 200 ArgAlaSerProAlaGlySerGlyAlaArgProGlnGlnSerThrAlaAlaGlnAla 218
Db 390 TCTTCTCAGCAGAGCTCTCTCTCTCAGCAGAGTCTCTCTCTCTCTCTCTCTCTCTCT 449

QY 219 -----ProGluAlaGlnAlaSerLeuProSerAlaSerSerGlnValAlaSer 235
Db 450 TCCCTCTCTCAGCAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 509

RESULT 5
US-09-864-761-15029
; Sequence 15029, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: A60MCA-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15029
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z99716.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
US-09-864-761-15029

Alignment Scores:
Pred. No.: 3.48e-15 Length: 471
Score: 245.00 Matches: 43
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.16% Indels: 0
DB: 9 Gaps: 0

US-09-864-291-12 (1-235) x US-09-864-761-15029 (1-471)
QY 98 Val11eVal1TyG1yAlaProProAlaG1yTyG1yAlaProProG1yTyG1yAla 117
DB 342 GTTATGCTATAGAGCCCACTGCGAGATATGAGCCCACTCCCGATACGAGCC 401
QY 118 ProProAlaG1yTyG1yAlaG1yAlaG1yAlaG1yAlaG1yAlaG1yAlaG1yAla 137
DB 402 CCACTGCGAGATATGAGCCCACTGCGAGATATGAGCCCACTCCCGATACGAGCC 461
QY 138 AlaSerPro 140
DB 462 GCCTCACCCT 470

RESULT 6
US-10-029-386-24981/c
; Sequence 24981, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
```

```

; SEQ ID NO 24981
; LENGTH: 511
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z97205.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
US-10-029-386-24981

Alignment Scores:
Pred. No.: 2.61e-13 Length: 511
Score: 226.00 Matches: 49
Percent Similarity: 45.87% Conservative: 1
Best Local Similarity: 44.95% Mismatches: 59
Query Match: 17.67% Indels: 0
DB: 13 Gaps: 0

US-09-864-291-12 (1-235) x US-10-029-386-24981 (1-511)
QY 102 G1yAlaProProAlaG1yTyG1yAlaProProG1yTyG1yAlaProProAlaG1y 121
DB 499 GGAAGAAGAGCTGCGAGGTGAGAGGAGAGCTGCGAGGTGAGAGGAGAGCTGCGAGGT 440
QY 122 TyG1yAlaG1yAlaG1yAlaG1yAlaG1yAlaG1yAlaG1yAlaG1yAlaG1yAla 141
DB 439 GAGGAGAAAGAGCTGCGAGGTGAGAGGAGAGCTGCGAGGTGAGAGGAGAGCTGCGA 380
QY 142 ArgTyG1yAlaProProLeuG1yTyG1yAlaProProAlaG1yTyG1yAlaProPro 161
DB 379 GGTGAGAGAGAGAGCTGCGAGGTGAGAGGAGAGCTGCGAGGTGAGAGGAGAGCTG 320
QY 162 LeuG1yTyG1yAlaProProLeuG1yTyG1yAlaProProLeuG1yTyG1yAlaPro 181
DB 319 GCAAGTGAAGAGAGAGCTGCGAGGTGAGAGGAGAGCTGCGAGGTGAGAGGAGAG 260
QY 182 ProLeuG1yTyG1yAlaProProAlaG1yAlaG1yAlaG1yAlaG1yAlaG1yAla 201
DB 259 CCTGAGGTGAGAGAGAGCTGCGAGGTGAGAGGAGAGCTGCGAGGTGAGAGGAG 200
QY 202 SerProAlaG1ySerG1yAlaArgPro 210
DB 199 GAGCTGCGAGGTGAGAGGAGAGCT 173

RESULT 7
US-10-029-386-11265
; Sequence 11265, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: ABOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11265
; LENGTH: 541
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO Z97205.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
; OTHER INFORMATION: SWISSPROT HIT: P22793, EVALU8 1.00e-04
US-10-029-386-11265

Alignment Scores:
Pred. No.: 7.82e-12 Length: 541
Score: 211.00 Matches: 53
Percent Similarity: 45.86% Conservative: 8
Best Local Similarity: 39.85% Mismatches: 48
```



```

FILE REFERENCE: ABOVICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 22839
LENGTH: 866
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005630.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: SWISSPROT HIT: Q08379, EVALUE 4.00e-10
OTHER INFORMATION: NT HIT: g1161055, EVALUE 4.00e-95
OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALUE 0.00e+00
US-10-029-386-22839

Alignment Scores:
Pred. No.: 7,91e-10 Length: 866
Score: 192.50 Matches: 45
Percent Similarity: 47.59% Conservative: 24
Best Local Similarity: 31.03% Mismatches: 73
Query Match: 15.05% Indels: 3
DB: 13 Gaps: 3

US-09-864-291-12 (1-235) x US-10-029-386-22839 (1-866)
QY 91 ThrProGlnMetProCysSerValIleValIYrGlyAlaProProAlaGlyYrGlyAla 110
DB 600 TCTCCGCGCCCTGAGAGCTTCTCTGTTCAACAGCCTCTCTCTGTTCAATAGCCTCT 541
QY 111 ProProProGlyYrGlyAlaProProAlaGlyYrGlyAlaGlnProAlaGlyAla 130
DB 540 CTTCTCTGTTCAACAGCCTCTCTCTCTGTTCAAGTACGCTCTCTCTCTGTTCAACAGCC 481
QY 131 GlyProProValGlyYrArgAlaSerProValArgYrGlyAlaProProAlaGlyYr 150
DB 480 TCTCCCTCTGTTCAAGTACGCTCTCTCTCTGTTCAACAGCCTCTCTCTCTGTTCAAGTA 421
QY 151 GlyAlaProProAlaGlyYrGlyAlaProProAlaGlyYrGlyAlaProProAlaGly 170
DB 420 GCTCTCTCTCTGTTCAACAGCCTCTCTCTCTGTTCAAGTACGCTCTCTCTCTCTGTTCA 361
QY 171 YrGlyYrThProProAlaGlyYrGlyAlaProProAlaGlyYrGlyAlaProProAla 190
DB 360 GTAGCCTCTCTCTCTGTTCAAGTACGCTCTCTCTCTCTGTTCAAGTACGCTCTCTCTCT 301
QY 191 GlyAlaGlnGlyProProAlaGlyYrArgAlaSerProAlaGlySerGlyAlaArgPro 210
DB 300 CAGGTACGCTCTCTCTCTGTTCAAGTACGCTCTCTCTCTCTGTTCAAGTACGCTCTCTCT 244
QY 211 GlnGlnSerThAlaAlaGlnAlaProGlnAlaGlnAlaSerLeuPro--SerAlaSer 229
DB 243 CCTGTTCAACAGCCTCT--CCTCTGTTCAAGTACGCTCTCTCTCTCTGTTCAAGTACG 187
QY 230 SerSerGlnValHis 234
DB 186 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 172

RESULT 12
US-09-764-891-1356
Sequence 1356, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231

```

```

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1356
LENGTH: 321
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (316)
OTHER INFORMATION: n equals a,t,g, or c
LOCATION: (321)
OTHER INFORMATION: (321)
US-09-764-891-1356

Alignment Scores:
Pred. No.: 3.09e-10 Length: 321
Score: 192.00 Matches: 39
Percent Similarity: 95.12% Conservative: 0
Best Local Similarity: 15.12% Mismatches: 1
Query Match: 15.01% Indels: 1
DB: 11 Gaps: 0

US-09-864-291-12 (1-235) x US-09-764-891-1356 (1-321)
QY 1 MetProPheApleMetThrasnLeuThrValGlnGlnProValPheAlaAsnPro 20
DB 195 ATGCCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 254
QY 21 IleYsGlyThrIleGlnAlaAlaProYrGlyGlyYrGlyAlaThrPhe 40
DB 255 ATTAAGGACATTATTACGAGCTCTCATATGAGCTCTGGAAGACAAAGCTACTTTA 314
QY 40 P 40
DB 315 A 315

RESULT 13
US-09-918-995-33536
Sequence 33536, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 33536
LENGTH: 432
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(432)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

Alignment Scores:
Pred. No.: 5.29e-10 Length: 432
Score: 191.00 Matches: 36
Percent Similarity: 72.31% Conservative: 11
Best Local Similarity: 55.38% Mismatches: 18
Query Match: 14.93% Indels: 0
DB: 11 Gaps: 0

US-09-864-291-12 (1-235) x US-09-918-995-33536 (1-432)
QY 1 MetProPheApleMetThrasnLeuThrValGlnGlnProValPheAlaAsnPro 20
DB 195 ATGCCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 254

```

Db 231 ATGCATTATCTCATGAAAGCTGTGATCAAGAGCCCGATTGTGCAAACTAC 230
Qy 21 ILLEAGLYTHRIIEGINAIAAIPROTYGILYITPGLUGLYGINAIASTRPHELYE 40
Db 291 ATCAAGGGAACAGAGGGGCGAAGCGGAGGTGGCTGGAGAGCTCTGCTTCTTACAG 350
Qy 41 LeuValPheArgAnGlyAsePAlaIleGluPheAlaGlnIleuMetValIysAlaIleSer 60
Db 351 TTGACTTCAAGGAGGGGGCGCATTTGATTCGAGACAGGAGATGCTCCAGGTGGATCT 410
Qy 61 AlavaIAaArgly 65
Db 411 CAAGCTCCAGAGGT 425
RESULT 14
US-10-029-386-20220
Sequence 20220, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: AEWICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 20220
LENGTH: 780
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005263.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2
OTHER INFORMATION: EST HUMAN HIT: AL539431.1, EVALUATE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: P04280, EVALUATE 1.00e-04
OTHER INFORMATION: NT HIT: g115307287, EVALUATE 0.00e+00
US-10-029-386-20220
Alignment Scores:
Pred. No.: 3.76e-09 Length: 780
Score: 185.00 Matches: 58
Percent Similarity: 38.86% Conservative: 17
Best Local Similarity: 30.05% Mismatches: 67
Query Match: 14.46% Indels: 51
Gaps: 7
US-09-864-291-12 (1-235) x US-10-029-386-20220 (1-780)
Qy 92 ProGlnMetProCySerValIleValTyrgly-----AlaProProAlaGlyTyr 108
Db 180 CCCCAAGAGGCTCGCGCTCGAGGCGCCCGGAGCAACCCAGCTAACCCCGCAGCTCC 239
Qy 109 GlyAlaProProProGly--TyrGlyAlaProProAlaGlyTyrGlyAlaGlnProVal 127
Db 240 AGGGAGTCCACCCCGCGCCGAGGTGGTCATCCCTGTCATCTGGGTCATCCGCCAGC 299
Qy 128 GlyAenGluGlyProProValIGlyTzArgAlaSerProValArgTyrGlyAlaProPro 147
Db 300 TCTCGGGCTCCACCCCGCCAGCTCTGAGGTTCATCCCGACCCCTGGGTCACACCC 359
Qy 148 LeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaPro-----ProLeuGly----- 163
Db 360 AACCTGGGGTCCACCCCGCAGCTCTGAGTCCACCTCCAGCCCGCGGGGTTCACCC 419

Qy 164 -----TyrglyAlaPro----- 167
Db 420 ACCAGCCCCGAGTCCACCCACAGCCCTCGGGGTTCACCCACAGCCCGAGGGTCA 479
Qy 168 -----ProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAlaPro----- 181
Db 480 TCTCTCCCATCAGCGGGGGTTCACCCCGAGGCCCGGGGGTTCACCCCGAGCCCGCC 539
Qy 182 -----ProLeuGlyTyr 185
Db 540 CATTACCTCCAGGCCCCAGGGGTGACCCACAGCCCGAGGAGTCCATCCAGGCC 599
Qy 186 GlyAlaProProAlaGlyAenGluGlyProProAlaGlyTzArgAlaSerProAlaGly 205
Db 600 GGGGGTCCACCCCGCAACTCCCGGGGTTCATCGTGGCTCTGGGGTCCACCTCAGCC 659
Qy 206 SerGlyAlaArgProGlnIleuSerThraIaIa-----GlnAlaProGluAen 221
Db 660 TCCGGAGTTCACCCCTCAATCTCGGGGTGACCCCGCAACTCCCATCCCATAGCT 719
Qy 222 GluAla-SerLeuProSerAlaSerSerGlnVal 233
Db 720 GAGGCCCCCATCTCCCTCCGAAAGCCCGAGGAACATA 756
RESULT 15
US-09-938-842A-2004
Sequence 2004, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Kreps, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
TITLE OF INVENTION: SAME, AND METHODS OF USE
FILE REFERENCE: SCRIPI300-3
CURRENT APPLICATION NUMBER: US/09/938,842A
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: US 60/227,866
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: US 60/264,647
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/300,111
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 5379
SEQ ID NO 2004
LENGTH: 2016
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-938-842A-2004
Alignment Scores:
Pred. No.: 1.03e-08 Length: 2016
Score: 185.00 Matches: 47
Percent Similarity: 41.35% Conservative: 8
Best Local Similarity: 35.34% Mismatches: 40
Query Match: 14.46% Indels: 38
Gaps: 6
US-09-864-291-12 (1-235) x US-09-938-842A-2004 (1-2016)
Qy 101 TyrglyAlaProProAlaGlyTyrGlyAlaProProProGlyTyrGlyAla----- 117
Db 355 TTGGGGGTTCAGAGGAATAAT-----CCACATATTAATTGGGTCATCGATGAG 405
Qy 118 -----ProProAlaGlyTyrGlyAlaGlnProValIGlyAen 129
Db 406 TTAAGGGTTCGAGATGAATCCCGCGCTTCACTGATGCTGCTTATTAACA 465
Qy 130 GlnGlyProProValIGlyTzArgAlaSerProValArg-----TyrGly 144
Db 466 -----CCACAGTGAATCAT-----CCACGTTAAGACAGCTCAGGGGGCTATGGA 513

Qy 145 AlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProLeuGlyTyr 164
Db 514 GTACCTCCAGCTGGCTATGAGATCAAGCTGGTTATGAGATCAAGCGGTTAT 573
Qy 165 GlyAlaProProLeuGlyTyrGlyThrProProLeuGlyTyrGlyAlaProProLeuGly 184
Db 574 GGAATACCTCAAGCTGGCTATGAGATCAAGCGGTTATGAGATCAAGCTCAAGTGGGC 633
Qy 185 TyrGlyAlaProProAlaGlyAangGlyProProAlaGlyTyrArgAlaSerProAla 204
Db 634 TAAGA-----ATGCCAGT 648
Qy 205 GlySerGlyAlaArgProGlnGlySerThrAlaAlaGln 217
Db 649 GGTTCCTCGAAGAGCTTGAATGAAGCAATGGCACTGAG 687

Search completed: December 17, 2003, 01:19:21
Job time : 15959.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 17:31:55 (Search time 1518.07 Seconds)

(without alignments)
3762.388 Million cell updates/sec

Title: US-09-864-291-12
Perfect score: 1279
Sequence: 1 MPDLMNTLVTEQVVFANP.....AQAPENASLPSSASQVHS 235

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Pgapop 6.0, Pgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPRO.spool/p/US09864291/runat.15122003.160859.23318/app.query.fasta.1.846
-DB=EST -OPMT=fastgap -SUFFIX=est -MINMATCH=0.1 -LOOCL=0 -LOOEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USRS=US09864291 @CGN 1.1 3724 @runat.15122003.160859.23318 -NCPU=6 -ICPU=3
-NO MAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEFTIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_rtd:*
27: em_gss_vit:*
28: gb_gss1:*

29: gb_gss2:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1207	94.4	1079	12	BMS64167 BMS64167 AGENCOURT
2	755.5	53.0	1414	11	AK015863 AK015863 Mus muscu
3	677.5	59.0	775	10	BG722815 BG722815
4	625	48.9	663	10	BG699398 BG699398
5	511.5	40.0	717	14	BT715414 BT715414
6	468	36.6	701	10	BG701881 BG701881
7	350	27.4	488	28	AQ997933 AQ997933
8	347.5	27.2	814	13	BU961805 BU961805
9	340.5	26.6	931	13	BU914288 BU914288
10	338.5	26.5	744	13	BU230211 BU230211
11	338	26.4	932	13	BU107273 BU107273
12	323.5	25.3	892	14	CD301781 CD301781
13	319	24.9	816	13	BU916890 BU916890
14	318	24.9	868	14	CD255258 CD255258
15	312	24.4	862	13	BQ735296 BQ735296
16	311.5	24.4	426	28	AQ997939 AQ997939
17	306	23.9	636	12	BM490729 BM490729
18	306	23.9	643	9	AL868685 AL868685
19	305	23.8	723	9	AJ452591 AJ452591
20	299.5	23.4	719	14	CB593075 CB593075
21	294	23.0	700	13	BU284977 BU284977
22	290.5	22.7	655	9	AL898723 AL898723
23	290.5	22.7	661	9	AL972453 AL972453
24	290.5	22.7	663	9	AL896682 AL896682
25	290	22.7	647	9	AL849651 AL849651
26	287	22.4	618	13	BU806987 BU806987
27	287	22.4	737	13	BX082055 BX082055
28	285	22.3	663	9	AL881810 AL881810
29	282.5	22.1	594	12	BM190972 BM190972
30	282.5	22.1	735	14	CA364472 CA364472
31	282	22.0	607	12	BM190975 BM190975
32	281	22.0	635	9	AL889954 AL889954
33	280.5	21.9	568	12	BT390687 BT390687
34	279	21.8	1004	13	BQ217128 BQ217128
35	279	21.8	1120	14	CD499460 CD499460
36	277.5	21.7	829	13	BU413257 BU413257
37	273.5	21.4	584	12	BJ501828 BJ501828
38	272	21.3	1028	13	BU541892 BU541892
39	271	21.2	747	13	BQ229609 BQ229609
40	270.5	21.1	922	13	BQ645308 BQ645308
41	269.5	21.1	718	14	CB591312 CB591312
42	269.5	21.1	776	13	BU846529 BU846529
43	269.5	21.1	878	13	BQ723331 BQ723331
44	269.5	21.1	913	13	BQ718987 BQ718987
45	269.5	21.1	937	13	BU174627 BU174627

ALIGNMENTS

RESULT 1
BMS64167
LOCUS BMS64167 1079 bp mRNA EST 20-FEB-2002
DEFINITION AGENCOURT 6560149 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742129
5', mRNA sequence.
ACCESSION BMS64167
VERSION BMS64167.1 GI:18811738
KEYWORDS
SOURCE EST.
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1079)

AUTHORS NIH-MGC <http://mgc.nhl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@nsb-nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LHAM2759 row: m column: 10
 High quality sequence stop: 726.

FEATURES
 source
 1. 1079
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5742129"
 /issue_type="medulla"
 /lab_host="DH10B"
 /clone_1lb="NIH MGC 119"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI;
 Site 2: EcoRV (destroyed); RNA source: normal medulla from
 anonymous male age 27. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.3 kb, insert size range
 0.9-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 013. Note:
 this is a NIH MGC Library."

BASE COUNT 267 a 297 c 268 g 247 t
ORIGIN

Alignment Scores:
 Pred. No.: 1.47e-70 Length: 1079
 Score: 1207.00 Matches: 229
 Percent Similarity: 97.06% Conservative: 2
 Best Local Similarity: 96.22% Mismatches: 4
 Query Match: 94.37% Indels: 3
 DB: 12 Gaps: 0

US-09-864-291-12 (1-235) x BM564167 (1-1079)

Qy 1 MeProheapleuMeTrrAsnLeuThVaIgInpProValPheAlaAAspHe 20
Db 269 ATGCCATTGATCTGATGACGAACTCTCACTTTAAACACGATTTCTGCAACTTC 328
Qy 21 TleYsglyThrTleglnAlaAProTyrglygTyrglyglnAlaArhPheLys 40
Db 329 ATTAAAGGAATATTCAAGGAGCTCCATATGCTGGCGGAGGACAAAGCTACTTTAAA 388
Qy 41 LeuValPheArGaAnglyAspAlaTleglnPheAlaGlnLeuMetVallyPheAlaSer 60
Db 389 TTACTCTTCAGAAATGAGGCGCATTTGAATTTCCAGGTGAAGGGAAGGCGCTCT 448
Qy 61 AlAValAlaArggLyPheProLeuArGThrLeuAsnAptTppheserSerMetGlyIle 80
Db 449 GCTCTGCGCGAGGATTTCCACTTAACCTTAATGACTGTTCAAGCTCTATGGAAAT 508
Qy 81 TyValIleThrglyglnGlyAsnMetCyArhThrProGlnMetProCySerValIleVal 100
Db 509 TATGTAATTACTGGGAGGAGGAAATATGCACTCAAGATGCCCTGTTCAGTTATGTC 568
Qy 101 TyrglyAlaProProAlaGlyTyrglyAlaProProProGlyTyrglyAlaProProAla 120
Db 569 TATGAGGCCCACTCGAGATATGAGCCCACTCCGATATGAGAGCCCACTCGCA 628
Qy 121 GlyTyrglyAlaGlnProValGlyAsnGlnGlyProProValGlyTyrglyAlaSerPro 140
Db 629 GGAATATGAGCCCACTCGATATGAGAGCCCACTCGATATGAGAGCCCACTCGCACT 688

Qy 141 VALATGTYTGYVALAProProLeuGlyTYTGYVALAProProAlaGlyTYTGYVALA-Pr 160
Db 689 GTGCATATGAGGCCCACTTGTGATAGGAGGCCCACTCGAGATATGAGGCCCC 748
Qy 160 oProLeuGlyTYTGYVALAProProLeuGlyTYTGYThrProProLeuGlyTYTGYAl 180
Db 749 ACCTCTAGATATGAGGCCCACTTGTGATATGAGGCCCACTCTCGATATATGAGAC 808
Qy 180 aProProLeuGlyTYTGYVALAProProAlaGlyAngGlnGlyProProAlaGlyTYTTr 200
Db 809 CCACCTCTCGATATGAGGCCCACTCGAGAAATGAGGCCCGCTCGGAGATACG 868
Qy 200 gAlaSerProAlaGlySerGlyAlaArGProGlnGlnSerThraAla-AlaGlnAlaProG 220
Db 869 AGCTCACTGCTCGATATGAGGCCCACTCGCAATATACGAGGCCCAAGCTCTCT 928
Qy 220 lAAsnGlu-AlaSerLeuProSerAlaSerSerSerGlnValAlaSer 235
Db 929 GAAAGAGAGCTTCT 976

RESULT 2
AK015863
LOCUS
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length cDNA library, clone:4930521123 product:hypothetical protein, full insert sequence.

ACCESSION AK015863
VERSION AK015863.1 GI:12854363
KEYWORDS HTC, CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Db 991 GCCACCTCTGTACACAGAGACCTCCACCTGCTGGCTGTGACAGACCATCCCATGCTC 1050

Qy 214 ThrAlaAlaGlnAlaProGluAaGlnAlaSerLeuProSerAlaSerSerGlnAla 233
|||||
Db 1051 GTGGCAGTCCAGATCTGAATTCACAGATCTTTTCCCTCACTGCTCCTCACAAGTC 1110
|||

Qy 234 HIsSer 235
|||||

Db 1111 CACTCT 1116

RESULT 3
LOCUS BG722815 775 bp mRNA linear EST 08-MAY-2001
DEFINITION 602695278F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5',
mRNA sequence.
ACCESSION BG722815
VERSION BG722815.1 GI:14002002
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 775)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
UNPUBLISHED
AUTHORS Contact: Robert Strausberg, Ph.D.
TITLES Email: cgabs-remail.nih.gov
JOURNAL Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
COMMENT CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshlyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM10743 row: 1 column: 20
High quality sequence stop: 651.
Location/Qualifiers
FEATURES
source 1..775
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4827427"
/lab_host="DH10B"
/clone_1lb="NIH_MGC_97"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVA-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASIS COUNT 191 a 212 c 193 g 179 t

ORIGIN

Alignment Scores:
Pred. No.: 1,01e-35 Length: 775
Score: 677.50 Matches: 144
Percent Similarity: 88.69% Conservativeness: 5
Best Local Similarity: 85.71% Mismatches: 13
Query Match: 52.97% Indels: 7
DB: 10 Gaps: 1

US-09-864-291-12 (1-235) x BG722815 (1-775)

Qy 1 MetProPheAspLeuMetTrpAsnLeuThrValGlnGlnProValIleAlaAspPhe 20
|||||

Db 276 ATGCCATTGATTCATGACGACACCTCAGCTTGAACACACAGATTTGCTGCAAACTTC 335
|||||

Qy 21 IlelysglyThrIleGlnAlaAlaProTyrlglylrrpGluGlnAlaThrPheLys 40
|||||

Db 336 ATTAAGGAACTATTCAGACAGCTCATATGCTGTGGAAAGACAACTCTTTTAA 395
|||||

Qy 41 LeuValPheArgAsnGlyAaAlaIleGlnPheAlaGlnLeuMetValIlyAlaIleSer 60
|||||

Db 396 TTAGCTTCAGAAATGAGAGTGCATTTGAATTTGGCAGATTATGATGTAAGCTGCTCT 455
|||||

Qy 61 AlAValAlAArgGlyPheProLeuAlaGlyThrLeuAen-AapTrpPheSerSerMetGly11 80
|||||

Db 456 GCTGTGCTCCGAGAGATTCACCTTAGAACCTTAACTGACTGATGCTATAGGAAT 515
|||||

Qy 80 eTyValIleThrGlyGlyGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
|||||

Db 516 TTATGTAACTTACTGGGAGAAAGGAAATATGTGACTCAGACAAATGCTTTTCACTATCT 575
|||||

Qy 100 TTTGTAlaAlaProProAlaGly-TyTGTAlaAlaProProGlyTyTGTAlaAlaProPro 120
|||||

Db 576 CTATGAGGCCCAACATGACGATATATGAGGCCCACTCCCGATATGAGAGC-CCAACTG 634
|||||

Qy 120 TAgTGTGTGTAlaGlnProValGlyAaGlnGlyProProValGlyTyTATGAlaSer- 139
|||||

Db 635 CAGGATATGAGACCCCAACCCGTAGAAACGAGCGCCGCTGTGGATTCAGAGCTCTCA 694
|||||

Qy 140 ProValArgTyr-GlyAlaProProLeuGly-TyTGTAlaAlaProProAlaGlyTyr---G 158
|||||

Db 695 CCTGTCCAGCAATGAAAGCCCACTCTCGGCAATAGCCAGCCCACTGCAAGCATAG 754
|||||

Qy 158 TAlaAlaProProLeu 162
|||||

Db 755 AAGCCCACTCTCA 768
|||||

RESULT 4
LOCUS BG699398 663 bp mRNA linear EST 07-MAY-2001
DEFINITION 602679081F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5',
mRNA sequence.
ACCESSION BG699398
VERSION BG699398.1 GI:13967653
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 663)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
UNPUBLISHED
AUTHORS Contact: Robert Strausberg, Ph.D.
TITLES Email: cgabs-remail.nih.gov
JOURNAL Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
COMMENT CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshlyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM10703 row: a column: 21
High quality sequence stop: 660.
Location/Qualifiers
FEATURES
source 1..663
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4811804"
/issue_type="hnpocampus"
/lab_host="DH10B"
/clone_1lb="NIH_MGC_95"
/note="Organ: Brain; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVA-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched


```

Db      677 AGAGCTCACC-----GGCGCCCA 697

RESULT 7
LOCUS   AQ97933/c                               488 bp    DNA        linear    GSS 24-FEB-2000
DEFINITION RPCI-23-384A20.TU RPCI-23 Mus musculus genomic clone RPCI-23-384A20
ACCESSION AQ97933
VERSION   AQ97933.1 GI:7072953
KEYWORDS  GSS.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
BUTRYOTA; Mearnsia; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 488)
Zhao,S., Nieman,W., Feldblum,T., Malek,J., Shatman,S., Akintiret
,B., Levine,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished
TITLE     Other_GSSs: RPCI-23-384A20.TU
JOURNAL   Contact: Shaying Zhao
COMMENT   Department of Bukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: szhao@tigr.org
          Clones are derived from the mouse BAC library RPCI-23. For BAC
          library availability, please contact Pieter de Jong
          (pietere@jlong.med.buffalo.edu). Clones may be purchased from
          BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
          or from Resea ch Genetics (into@resgen.com). BAC end page:
          http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html
          Plate: 384 row: A column: 20
          Seq primer: SP6
          Class: BAC ends.
FEATURES             Location/Qualifiers
     source            1..488
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-384A20"
                     /sex="Female"
                     /lab_host="DH10B"
                     /clone_1lb="RPCI-23"
                     /note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
                     Score1; Site_2: Score1; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and partially digested
                     with a combination of EcoRI and EcoRI Methyase. Size
                     selected DNA was cloned into the pBAC3.6 vector at the
                     EcoRI sites. The ligation products were transformed into
                     DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT  102 a       128 c       157 g       100 t         1 others
ORIGIN
Alignment Scores:
Pred. No.:           2 77e-14              Length:           488
Score:               350.00                Matches:           73
Percent Similarity: 52.26%                 Conservative:      8
Best Local Similarity: 47.10%               Mismatches:       34
Query Match:        27.37%                  Indels:          40
DB:                28                       Gaps:            4

US-09-864-291-12 (1-235) x AQ97933 (1-488)

QY      101  TTGTGTAATPProFAlAGlTYTGTGTAATPProFAlTYTGTGTAATPProFAlA 120
DB      466  TATTGGGCTCACCGTGGAATACGAGTCATCTGCAGGTTATGAGCCCATCTCG 407
QY      121  GTTGTGTGTAATPProFAlAGlTYGANGlUGlTPProFAlAGlTYTGTGTAATPPro 140

```

Dn		406	GAGTACGAGAGNCCGCCAGTGGGTTATGTAGACCCTCCTCCCGATTACGATGTCCTGCCT	347
Oy		141	-----ValArgTYrGYAlaPro-----	Proleu 148 :::
Dn		346	CCTGATATACGAAGCTGTGATATAGCTTGCTCCTCCTCTATATATAGTACTACCCCAATG	287
Oy		149	GLYTYrGYAlaPProProAlaGLYTYrGYAlaAProProLeuGLYTYrGYAlaAProPro	168
Dn		286	GAGTATAGAGTTCACCTCCTGGATATAGACCTCACCTGTGATGATNGGATCCCACT	227
Oy		169	LeuGLYTYr-----	171
Dn		226	CTGGATATGTAGAAGCCCCCATATGATGATAGAGCCCACTCCCATGATATAAAAACTACA	167
Oy		172	-----GLYThrProProLeuGLYTYrGYAlaAProProLeuGLYTYrGYAla	187
Dn		166	CTATAGGATCCGGTCCCACCTCCCATATGAAAGCCCCGCTATGGGGTATAGAACT	107
Oy		188	ProProAlaGLYANGluGLYProProAlaGLYTYrArgAlaSer-----	202
Dn		106	CCTCTCTTAGAGGCGTGAAGTCCATACCTCTGGATATCAAGGGCCACAGTCTGTAGACACAGAA	47
Oy		203	-----ProAlaGLYserGLYAlaArgProGLInGLyserThrAla	215
Dn		46	GCTCCACCTGCTGGGTCTGAAGACAGAGCCATCCCATGTCTGTGCA	2
RESULT 8				
LOCUS	BUS961805	814 bp	mRNA	linear EST 21-OCT-2002
DEFINITION	AGENCOURT 10617160 NIH_MGC_169 Mus musculus cDNA clone			
VERSION	BU961805			
KEYWORDS	BU961805.1 GI:24191377			
SOURCE	EST.			
ORGANISM	Mus musculus (house mouse)			
REFERENCE	Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 814) NIH-MGC http://mgs.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D. Email: cgabds-remail.nih.gov Tissue Procurement: Dr. Jonathan Kuo, NIMH cDNA Library Preparation: Michael Brownstein Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.jnl.gov Plate: LNCM3080 row: 1 column: 06 High quality sequence stop: 504. Location/Qualifiers 1..814 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /lab_host="IMAGR:6742447" /clone_id="NIH_MGC_169" /note="Organ: Testicles; Vector: pDNR-LIB; Site 1: Self (ggccatgagcc); Site 2: Self (ggcgctcgcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCATGTATCACGACAGATGGCCATTAAGCCCGG-3' and 5'-ATTCTAAGAGCCAGAGCGGCCGACATG-dt(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC library."			
FEATURES	Source			
COMMENT				
JOURNAL				
AUTHORS				
TITLE				
ABSTRACT				
KEYWORDS				
DESCRIPTORS				
TEXT				
FIGURE				
TABLE				
SEQUENCE				
OTHER				
REMARKS				
DATE				
ENTRY				
LAST EDITED				
BASE COUNT	207 a 238 c 177 g 190 t 2 others			

Alignment Scores:

Pred. No.: 6,61e-14 Length: 814
 Score: 347.50 Matches: 74
 Percent Similarity: 58.16% Conservative: 8
 Best Local Similarity: 52.48% Mismatches: 38
 Query Match: 27.17% Indels: 21
 DB: 13 Gaps: 3

US-09-864-291-12 (1-235) x BU914288 (1-814)

QY 102 GYLAAPProProlaaglyTYrglyYLAAPProProglYTYrglyYLAAPProProlaagly 121
 DB 3 GGCTCCCTCTCTCTCTATATGTATGCTACCCCAATGGCTATGAGTTCACCTCTCGGA 62
 QY 122 TYrglyYLAAGInProValGlyAAGInGlyProProValGlyTYrArgAlaSerProVal 141
 DB 63 TAT-----GACCTTCA-----CCTGTG 80
 QY 142 ArgTYrglyYLAAPProProleuglyTYrglyYLAAPProProlaaglyTYrglyYLAAPPro 161
 DB 81 AGATACCGATCCCACTCTCTGATATGAAAGCCCACTATGAGTATGAGCCCAACT 140
 QY 162 LeuglyTYrglyYLAAPProProleuglyTYrglyYrProProleuglyTYrglyYLAAPro 181
 DB 141 CCCAGATACCGAACTACACCTATGGATCCGGTCTCACCTCCAGATATGAAAGCCCG 200
 QY 182 ProleuglyTYrglyYLAAPProProlaaglyYAsnGInGlyProProlaaglyTYrArgAla 201
 DB 201 CCAATGGGATATGAGTACTCTCTCTGAGGCGTGAAGTCATPACTCTTGATCAAGGCC 260
 QY 202 Ser-----ProAlaGlySerGlyYLAAPProProInGlySerThr 214
 DB 261 ACCTCTTACACAGAAAGCTCCACCTCTGCTGCTTAAGCAAGCCATCCCATGTCTGTG 320
 QY 215 AlaAlaGlnAlaProGluAAGInGlyAAserLeuProSerAlaSerSerGlnValHis 234
 DB 321 GCATGCCAGATCTGAATTCAGAGCATCTTTCCCTTACCTGCTCTCAAGTCCAC 380
 QY 235 Ser 235
 DB 381 TCT 383

RESULT 9

BU914288 931 bp mRNA linear EST 17-OCT-2002
 LOCUS BU914288 931 bp mRNA linear EST 17-OCT-2002
 DEFINITION AGENCOURT_10492155 NICHID_XGC_OO1 Xenopus laevis cDNA clone
 INAME:6640777 5', mRNA sequence.

ACCESSION BU914288
 VERSION BU914288.1 GI:24096202
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 Xenopodinae; Xenopus.

REFERENCE

1 (bases 1 to 931)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov

JOURNAL

TISSUE Procurement: Martha Redbert, Steven L. Klein, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov.
 Plate: LNL4223 row: 1 column: 01
 High quality sequence stop: 654.

FEATURES

Location/Qualifiers

Source

1..931
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:6640777"
 /issue_type="cocytes"
 /lab_host="DH10B (phage-resistant)"
 /clone_11b="NICHID_XGC_OO1"
 /note="Vector: PCMV-SF0R6; Site 1: NotI; Site 2: SalI;
 cloned unidirectionally. Primer: oligo dt. Average insert
 size 2.2 kb. Constructed by Life Technologies."

BASE COUNT 254 a 256 c 201 g 220 t
 ORIGIN

Alignment Scores:

Pred. No.: 2.18e-13 Length: 931
 Score: 340.50 Matches: 90
 Percent Similarity: 45.89% Conservative: 16
 Best Local Similarity: 38.96% Mismatches: 67
 Query Match: 26.62% Indels: 58
 DB: 13 Gaps: 11

US-09-864-291-12 (1-235) x BU914288 (1-931)

QY 1 MetProPheAapleuMetThrAsnLeuThrValGluGlnProValPheAlaAsnPro 20
 DB 275 ATGCCATTCACATGATGAAAGATGCTCATAGAAACAGCCAGTGTCTCTGCCAATAC 334
 QY 21 IleValGlyThrIleGlnAlaAlaProTYrglyTYrGInGlyGlnAlaThrPheVal 40
 DB 335 ATCAAGGAAACCATCATGCTGAACTGAGGTGGCTGAGAGGCTCAACCTCTTTTAA 394
 QY 41 LeuValPheArgAsnGlyAspAlaIleGlnPheAlaGlnLeuMetValYsAlaSer 60
 DB 395 ATTACATTATATAGTGAAGTGTCTATGATTTGACAAATCATGTTCAAAATGCAACT 454
 QY 61 AlaValAlaArgGlyPheProLeuAryThrLeuAsnAspTrpPheSerMetGlyLe 80
 DB 455 TGTGCTTCCAGA----- 466
 QY 81 TYrValIleThrGlyGlnGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
 DB 467 -----ACCCCTCTGCTTCATGCTCATATGGA 496
 QY 101 TYrGlyYLAAPProProlaaglyTYr-----GlyYLAAPProProProGlyTYrgly 116
 DB 497 TACACACCTGCTCCAGAGCTATGCAATGGCGGATACCTCCAGCTCAAGAAATATAT 556
 QY 117 AlaProProlaaglyTYrglyYLAAGInProValGlyAAsnGInGlyProProValGlyTYr 136
 DB 557 ACACCCCA-----CCCAACCATCAGGC----- 580
 QY 137 ArgAlaSerProValArgTYrglyYLAAPProleu---GlyTYrglyYLAAPProAla 155
 DB 581 -----CCTTATCCATATGAGCCCTGCTATGAAATGATATGAGACCACTCCAAAT 631
 QY 156 GlyTYrglyY-----AlaProProleuglyTYrglyYrProProleuglyTYrgly 179
 DB 632 CCAATAGATACCCCGATATGCTATCCACAGAGACCTCGCATATACACCCCTCTGAA 691
 DB 692 ATGAATCCATTTATGAGCTCTCTCCACACCTTATCTGAGACCCCTTAAAT---GGG 748
 QY 180 AlaProProleuglyTYrglyYLAAPro-----ProAlaGly-----AsnGInGlyPro 195
 DB 749 ACCCCCTACAAAGAAACCCCACTCTGCTCCCACTGATGCAATGATGCTGGTATG 808
 QY 196 ProAlaGlyTYrArgAlaSerProAlaGlySer 206
 DB 809 CTGAGAGTACCAAGCTGCAAAAGCTGATCC 841

RESULT 10

BU230211
 LOCUS BU230211 744 bp mRNA linear EST 26-NOV-2002
 DEFINITION 603398308F1 CSBQCHN23 Gallus gallus cDNA clone CHEST28901 5', mRNA
 sequence.
 ACCESSION BU230211
 VERSION BU230211.1 GI:25470774
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Bularicta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 744)
 Boardman, P.B., Sanz-Bzquez, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.,
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 JOURNAL PUBMED
 MEDLINE 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
)
 PO Box 88, Manchester, M60 10D, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 FEATURES
 source
 1..744
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST28901"
 /dev_stage="22"
 /lab_host="DH10B"
 /clone_lib="CSBQCHN23"
 /note="Organ: heads; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunt-ended, ligated to NotI adaptors, digested with EcoRI
 , size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."
 BASE COUNT 164 a 225 c 179 g 176 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,386-13 Length: 744
 Score: 338.50 Matches: 81
 Percent Similarity: 44.93% Conservative: 12
 Best Local Similarity: 39.13% Mismatches: 57
 Query Match: 26.47% Indels: 57
 DB: 13 Gaps: 6
 US-09-864-291-12 (1-235) * BU230211 (1-744)
 Oy 1 MetProPheAspLeuMetThrAsnLeuThrValGluGlnProValPheAlaAlaAspHe 20
 Db 45 ATGCCGTTTATTATTTGTAAGAGCGTCTATTAGACGCGCTTTCTGCTGCTAATTAC 104
 Oy 21 TlelyGgylThrlGlnAlaAlaProTyrglyGlyTTPglGlyGlnAlaThrPheIys 40
 Db 105 ATCAAGAGACAGATTCAGGCTGAGGACAGGAGTGTGGTGAAGGCGACGCTTTAA 164
 Oy 41 LeuValPheArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetVallyAlaAlaIaSer 60

Db 165 CTGACTTTCACACACGAGGAGGACCATGAGTTTGACAGCTGAGTTCACAAAGTGGCTT 224
 Oy 61 AlAValAlaArgGlyPheProLeuArgThrLeuAsnAptTrpPheSerSerMetGlyIle 80
 Db 225 AGTCTTCAGATGAGATTCCTCTCAG----- 251
 Oy 81 TyrValIleThrGlyGlyGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
 Db 251 ----- 251
 Oy 101 TyrGlyAlaProProAlaGlyTyrGlyAlaProProGlyTyrGlyAlaProProAla 120
 Db 252 -----CCCCCTGCTATGATACACACTGTA 278
 Oy 121 GlyTyrGlyAlaGlnProValGlyAsnGlyGlyProProVal-----GlyTyrArgAla 138
 Db 279 CCTGAGGGATATGACCTGTC-----CCAGCTGCTCAGAGAGGTATGCACT 326
 Oy 139 SerProValArgTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGlyTyrGly 158
 Db 327 GCTCCAGAGAGGTATGACCTGCTCAGAGAGGTATGACCTGCTCAGAGAGGTATGCA 386
 Oy 159 AlaProProLeuGlyTyrGlyAlaPro-----ProLeuGlyTyrGly 172
 Db 387 CCTGCTCCGAGGAGGTATGCTCTCTCCACACCTCCAAATGCTTATCTTATGCA 446
 Oy 173 ThrProProLeu-----GlyTyrGlyAlaProProLeuGlyTyrGlyAlaProProAlaGly 191
 Db 447 CCACCTCCGATGAATGCTTACGAGACCTGCTCCAGCCCATGGGTATCTGTATGCCAG 506
 Oy 192 AsnGlyGly-----ProPro 196
 Db 507 AATCAGGTATTATACCAACG 527
 RESULT 11
 BU107273 932 bp mRNA linear EST 25-NOV-2002
 LOCUS BU107273
 DEFINITION 603111954F1 CSBQCHL12 Gallus gallus cDNA clone CHEST60k13 5', mRNA
 sequence.
 ACCESSION BU107273
 VERSION BU107273.1 GI:25309754
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Bularicta; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 932)
 Boardman, P.B., Sanz-Bzquez, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.,
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 JOURNAL PUBMED
 MEDLINE 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
)
 PO Box 88, Manchester, M60 10D, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 FEATURES
 source
 1..932
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST60K13"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_lib="CSBQCHL12"


```

ACCESSION      CD255258
VERSION        CD255258.1  GI:31015724
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis (African clawed frog)
Xenopus laevis
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
Xenopodinae; Xenopus.
1 (bases 1 to 868)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NCI
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cga@nci.nih.gov
Tissue Procurement:
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LIML4578 row: 9 column: 04
High quality sequence stop: 673.
Location/Qualifiers
1..868
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:6954461"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NICHX.XC.Bml"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC)
library."
BASE COUNT      241 a      244 c      189 g      193 t      1 others
ORIGIN
Alignment Scores:
Pred. No.:      6,24e-12      length:      868
Score:          318.00      Matches:      85
Percent Similarity: 44.74%      Conservative: 17
Best Local Similarity: 37.28%      Mismatches: 78
Query Match:    24.86%      Indels:      48
DB:            14      Gaps:        8
US-09-864-291-12 (1-235) x CD255258 (1-868)
QY      1 MetPropheAspLeuMetThrAsnLeuThrValGluGlnProValPheAlaIasnPhe 20
Db      253 ATGCATTCTACCTGATGTAAGAGATGCTCCATAGAACACGCCAGTCTTCTGCGCAACTAC 312
QY      21 IIElyRGlyThriIleGlnAlaIaProTyRGlyGlyTyrGluGlyGlnAlaIaThrPheIys 40
Db      313 ATCAAGGGAACCATCATGCTGAAACCTGAGATGGCTGGAGAGGTAAACCTCTTTTAA 372
QY      41 LeuValPheArgAsnGlyAspAlaIleGluPheAlaGlnLeuMetValIysAlaIaSer 60
Db      373 CTTACCTTTTAACAGTGGAGAGACTATTGAATTGGACAAATCATGTTCCAAATATGGCAACT 432
QY      61 AAlaValAlaArgGlyPheProLeuArgThrLeuAsnAspTrpPheSerSerMetGlyIle 80
Db      433 TCTGCTTCCAGAGCCCTCTCTGT----- 456
QY      81 TTTTAlIleThrdGlyGluGlyAsnMetCysThrProGlnMetProCysSerValIleVal 100
Db      456 ----- 456

```

QY		101	TyrglYAlaProProlaaglyTyrglYAlaProProloglyTyrglYAlaProProla	120
Dd		457	-----CCTAGTGGATATGTGATACACTCCCTCGACAGAGAATATGACCAAGC	507
QY		121	GlyTYrglYAlaGlnProProValGlyaBnglUglyProProValGlyTYrAgaLaserPro	140
Dd		508	GGGTACCCTTCACAGTACAGGAATATATACACCCCCAACCC-----CAACCATCAGGCCT	561
QY		141	ValArgTYrglYAlaProProleu----GlyTYrglYAlaProProalaglyTYrglY---	158
Dd		562	TATCATATGAGACCCCCTGTAATAATGAAATGATATGACCACTCCGAATGCTATGAGATAC	621
QY		159	-----AlaProProleuglyTYrglYAlaProProleuglyTYrglYThnProProleu	176
Dd		622	CCGATAGCTCCGCCACAGAGACCTGGATATATCACACCCCTCCTGAATGAATCCATT	681
QY		177	GlyTYrglYAlaProProleuglyTYrglYAlaProProalaglyaBnglUglyProPro	196
Dd		682	---TACATGCTCTCTCCACCACTTATCTCGAACCCCG-----TATATGAGACCCCCA	732
QY		197	AlagLYrArAgaLaserProalaglySerGlyAlaArgProGlngInuSerThnAlala	216
Dd		733	-----TATATGAGACCCGACCACTTGAGCT-----CCCTGATCATGATATCT	780
QY		217	GlnAlaProGluBnglualaser	224
Dd		781	GGTATGCTCGGAGGACGACCAAG	804
RESULT 15				
BQ735296				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
SOURCE				

Alignment Scores:

Pred. No.:	1.54e-11	Length:	862
Score:	312.00	Matches:	82
Percent Similarity:	43.04%	Conservative:	17
Best Local Similarity:	35.65%	Mismatches:	71
Query Match:	24.39%	Indels:	60
DB:	13	Gaps:	6

US-09-864-291-12 (1-235) x BQ735296 (1-862)

OY	1	MetProPheAspLeuMetThrIleuLeuThrValGlnGlnProValPheAlaAlaAspPhe	20
Db	224	ATGCCATTCTACTGTAAGAAAGAGTCTCATTAAGAACGCCAGTCTTCTCGCAACTAC	239
OY	21	IleYsgLYThrIleGlnAlaAlaProTYrgLYTYrPoluGLYglnAlaThrPheLys	40
Db	294	ATCAAGGGAACCATCATGCTCTGAACTGGAGAGTGGCTGGAGGGTAAACCTCTTTAA	353
OY	41	LeuValPheArgAsnGLYAspAlaIleGluPheAlGlnLeuMetVallybAlaAsp	60
Db	354	CTTACATTAAACAGTGGAGAGAGCTATTGAATTGGCAAAATCATGTTCAAAATGGCAACT	413
OY	61	AlaValAlaArgGLYpHeProLeuAGThrLeuAsnAspTrpPheSerMetGLYIle	80
Db	414	TCCTGCTTCAGAGGCCCTCTCTT-----	437
OY	81	TyrValIleThrGLYglnGLYAsnMetCysThrProGlnMetProCysSerValIleVal	100
Db	437	-----	437
OY	101	TyrGLYAlaProProAlaGLYTYrgLYAlaProProProGLYTYrgLYAlaProProAla	120
Db	438	-----CCTCAGCTGCATATGATATACACTCCTGCTGACAGAGATATGACCAAGC	488
OY	121	GLYTYrgLYAlaGlnProValGlsAsnGLYGLYProProValGLYTYrArgAlaSerPro	140
Db	489	GGGTACCTCCAGCTCCAGAAATATATACACCCCAACC-----CAACCATCAGGCCCT	542
OY	141	ValArgTYrgLYAlaProProLeu---GLYTYrgLYAlaProProAlaGLYTYrgLY---	158
Db	543	TATCATATATGAGACCCCTGCTATGAAATGATATGAGCAACCTCCGAATGCTATGGATAC	602
OY	159	-----AlaProProLeuGLYTYrgLYAlaProProLeuGLYTYrgLYThrProProLeu	176
Db	603	CCGTAATGCTCCGCCACACAGACCTGCGAGATATCAACCCCTCTGTAATGAATCCCAT	662
OY	177	GLYTYrgLYAlaProProLeuGLYTYrgLYAlaProProAla-----	190
Db	663	---TACATGGCTCTTCCACCACTTATCTCGAGCCCCCGTATATATGAGACCCCATATAA	719
OY	191	-----GLYAsnGLuGLYProProAlaGLY	198
Db	720	TGGAACCCAGACCTTCAGCTCCCTGTGATTCATGNAATACTGGTATGCTCGAGGCG	779
OY	199	TyrArgAlaSerProAlaGLYSerGLYAla	208
Db	780	AGCAAGCCACAGAACGACATCTCAAGTGA	809

Search completed: December 16, 2003, 20:45:15
Job time : 1526.57 secs

SUMMARIES

10999.596 Million cell updates/sec

Sequence: 1 cggcagcagggcggcagga.....gcataaaaaaaaaaaaaa 1413

Searched: 2888711 reqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenBank :

1:	gb_ba.*
2:	gb_hg.*
3:	gb_in.*
4:	gb_om.*
5:	gb_ov.*
6:	gb_pac.*
7:	gb_ph.*
8:	gb_pl.*
9:	gb_pt.*
10:	gb_to.*
11:	gb_sac.*
12:	gb_sy.*
13:	gb_un.*
14:	gb_vt.*
15:	em_ba.*
16:	em_fun.*
17:	em_hum.*
18:	em_in.*
19:	em_mn.*
20:	em_om.*
21:	em_ot.*
22:	em_ov.*
23:	em_pat.*
24:	em_ph.*
25:	em_pl.*
26:	em_to.*
27:	em_sac.*
28:	em_un.*
29:	em_vt.*
30:	em_hg_hum.*
31:	em_hg_inv.*
32:	em_hg_other.*
33:	em_hg_mus.*
34:	em_hg_pln.*
35:	em_hg_rod.*
36:	em_hg_mam.*
37:	em_hg_vrt.*
38:	em_gv.*
39:	em_hgo_hum.*
40:	em_hgo_mus.*
41:	em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	1413	100.0	1413	6	AX35658	AX35658 Sequence
2	592.4	41.9	2267	9	BC022546	BC022546 Homo sapi
3	587.2	41.6	2266	6	BC022549	BC022549 Homo sapi
4	435.4	30.8	1001	6	AX359665	AX359665 Sequence
5	347.2	24.6	179222	2	BX296515	BX296515 Suf scrof
6	347.2	24.6	232459	2	BX470149	BX470149 Dario rer
7	282.6	20.0	220895	9	HS250D10	Z59716 Human DNA B
8	223.2	15.8	129624	2	BX296540	BX296540 Suf scrof
9	188.8	13.4	168425	2	AC113593	AC113593 Mus muscu
10	182.6	13.4	224086	10	AC104325	AC104325 Mus muscu
11	172.6	12.2	255149	2	AC107527	AC107527 Rattus no
12	172.6	12.2	270171	2	AC132969	AC132969 Rattus no
13	128.4	9.1	793	10	AF439026	AF439026 Rattus no
14	128.4	9.1	1802	10	MP040826	U40826 Mus muscucu
15	122.8	8.7	222469	9	BX470149	BX470149 Dario rer
16	122	8.6	1691	9	HSU79458	U79458 Human MW do
17	122	8.6	1820	9	BC007452	BC007452 Homo sapi
18	122	8.6	1885	6	AR220846	AR220846 Sequence
19	122	8.6	1912	9	BC010616	BC010616 Homo sapi
20	121.6	8.6	2170	9	AB072784	AB072784 Macaca fa
21	120.4	8.5	2170	9	AK057881	AK057881 Homo sapi
22	116	8.2	1205	5	BC053232	BC053232 Dario rer
23	116	8.2	95108	2	AC127818	AC127818 Rattus no
24	110.2	7.8	106438	9	AL365445	AL365445 Human DNA
25	106.4	7.5	2055	14	AP40847	AP40847 Oryzine her
26	106.4	7.5	148418	9	AC02698	AC02698 Homo sapi
27	103.8	7.3	454	6	BD027759	BD027759 Sequence
28	102.8	7.3	246814	2	AC103482	AC103482 Rattus no
29	100	7.1	235381	2	AC103031	AC103031 Rattus no
30	96	6.8	247947	2	AC098163	AC098163 Rattus no
31	95.8	6.6	110000	2	AC118467_0	AC118467 Mus muscu
32	93.8	6.6	216292	2	AC100927	AC100927 Mus muscu
33	91.6	6.5	235972	2	AC095835	AC095835 Rattus no
34	91.6	6.5	241519	2	AC097689	AC097689 Rattus no
35	91.6	6.5	250902	2	AC131965	AC131965 Rattus no
36	91.6	6.5	251746	2	AC131411	AC131411 Rattus no
37	91.4	6.5	208880	2	AC128103	AC128103 Rattus no
38	90.6	6.4	230559	2	AC108250	AC108250 Rattus no
39	90.6	6.4	238377	2	AC121702	AC121702 Rattus no
40	90.4	6.4	210680	2	AC123724	AC123724 Mus muscu
41	90.4	6.4	212334	2	AC023070	AC023070 Mus muscu
42	90.4	6.4	315390	2	AC124097	AC124097 Mus muscu
43	90	6.4	262967	2	AC110392	AC110392 Rattus no
44	89.8	6.4	261914	2	AC122581	AC122581 Rattus no
45	88.8	6.3	88716	2	AC098644	AC098644 Takifugu

ALIGNMENTS

AX359658	AX359658	1413 bp	DNA	linear	PAT 13-FEB-2002
LOCUS	AX359658				
DEFINITION	Sequence 4 from Patent WO0190185.				
ACCESSION	AX359658				
VERSION	AX359658.1	GI:18675409			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	1				
TITLE	Okamoto, R. and Sultovskiy, P.				
JOURNAL	Pt22 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses thereof				
	Patent: WO 0190185-A 4 29-NOV-2001;				

QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES
UNIVERSITY (US)
Location/Qualifiers
.....

Source

252

```

CDs
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="mammalia"
36.977
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD2336.1"
/db_xref="GI:18675410"
/translation="MAVNSHTSESRGALLPSGSVYLKQEDVDLCTFLQKRVESYLNN
GTRKGLPLTSYRVVFPVTSILVNDPMLSFMPFELMSCTTBOTIPAPNTIKGTTQA
PGGMEQGVAFKLSFRKQGLAEQALWKAASAARGLPLGVNWPDTSLYITVPA
GAVCSQTPCPDAVIVIVGPPPGYITVQPBSTYTPRSGYACGGGAGAPPMWGAGAPP
VGVGVPGGAGVPPRGGYVPPGGYGAAPGGYGVPPGGYAPPGYAPPAQYVADAPPAQ
MEALPDAEAPASAGNTAASHRSMTAQCEITSLPTSSS"
BASE COUNT      377 a      363 c      369 g      304 t
ORIGIN

```

[illegible]

QY	721	GAGGCCCACTGGGGGGATATGAGATCCACCTGGGGGATATGAGGCCCACTGGGGGAT	780
Db	721	GAGGCCCACTGGGGGATATGAGATCCACCTGGGGGATATGAGGCCCACTGGGGGAT	780
QY	781	ATGAGATCCCACTGGGGGATATGATGCCCACTGGGGGATATGAGGCCCACTGGAG	840
Db	781	ATGAGATCCCACTGGGGGATATGATGCCCACTGGGGGATATGAGGCCCACTGGAG	840
QY	841	GATATGAGGCCCACTGAGTAATGAGCCCTACCCCTGCAATAGAGCTTCATCTG	900
Db	841	GATATGAGGCCCACTGAGTAATGAGCCCTACCCCTGCAATAGAGCTTCATCTG	900
QY	901	CTGAGAAATACAGCTGCTCTCACAAGATCTATGACAGTCAAGCAGAGAGACTTCTTCCCA	960
Db	901	CTGAGAAATACAGCTGCTCTCACAAGATCTATGACAGTCAAGCAGAGAGACTTCTTCCCA	960
QY	961	CTACTCATCTCTTAAAGTCAATTATACCACTTCTGAGAGTTAAACCTTGAAGCTCAC	1020
Db	961	CTACTCATCTCTTAAAGTCAATTATACCACTTCTGAGAGTTAAACCTTGAAGCTCAC	1020
QY	1021	AAGCAAGGGGACCCCTAAACTGAAGTCAAGTAAAGAGAGAGCCAGGTGCCAGTGG	1080
Db	1021	AAGCAAGGGGACCCCTAAACTGAAGTCAAGTAAAGAGAGAGCCAGGTGCCAGTGG	1080
QY	1081	TAGAGAGTGTTCGTGTGACAGCAAGTGGTCTGATTTTCTCCACACACTGTGAGGTCTGT	1140
Db	1081	TAGAGAGTGTTCGTGTGACAGCAAGTGGTCTGATTTTCTCCACACACTGTGAGGTCTGT	1140
QY	1141	GCCCTCAAAACAGATGAAGGTGAAGAGACGACTCTGTCTTCAAGAGAGAGAGAGTCTTGA	1200
Db	1141	GCCCTCAAAACAGATGAAGGTGAAGAGACGACTCTGTCTTCAAGAGAGAGAGAGTCTTGA	1200
QY	1201	AAACAGACTGCAACCACTAGAGAGAGAGATGTGAAGTGCACATAAACAGCTTGG	1260
Db	1201	AAACAGACTGCAACCACTAGAGAGAGAGATGTGAAGTGCACATAAACAGCTTGG	1260
QY	1261	GGATGAGACTGACTCTCTTTAGAAAACAGGCTCTTCCCTGCTTGAAGCTGAGAGAA	1320
Db	1261	GGATGAGACTGACTCTCTTTAGAAAACAGGCTCTTCCCTGCTTGAAGCTGAGAGAA	1320
QY	1321	AAGAGAAATCCTGGAACCAAGAGCTTAAAGGTCACTCTGCTTAAAGCCCTTCGATTAAAG	1380
Db	1321	AAGAGAAATCCTGGAACCAAGAGCTTAAAGGTCACTCTGCTTAAAGCCCTTCGATTAAAG	1380
QY	1381	CCTGCTTGCTGTGCATTAATAAAAAAAAAAAAA	1443
Db	1381	CCTGCTTGCTGTGCATTAATAAAAAAAAAAAAA	1443

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	AUTHORS	JOURNAL	REMARK	COMMENT	
BC022546	2267 bp mRNA linear	BC022546	BC022546	MGC.	Homo sapiens (human)	Homo sapiens	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	1 (baaes 1 to 2267)	Strausberg, R.	Submitted (01-FEB-2002)	National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	NIH-MGC Project URL: http://mgc.nci.nih.gov	Contact: MGC help desk Email: cgasbhe-remail.nih.gov Tissue Procurement: Miklos Palokovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA library arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdick@lln.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MOC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.linl.gov>
Series: IRAP Plate: 32 Row: 3 Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES

Location/Qualifiers

Bouice

```
source
1. .2267
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26816 IMAGE:4811804"
/cisue_type="Brain, hippocampus"
/clone_1ib="NIH MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
32. .961
DSD
```

CDS

BASE COUNT ORIGIN	709 a	477 c	469 g	612 t
----------------------	-------	-------	-------	-------

Query Match	41.9%	Score 592.4;	DB 9;	Length 2267;
Best Local Similarity	75.7%	Pred. No. 5.9e-136;		
Matches 796;	Conservative 0;	Mismatches 231;	Indels 25;	Gaps 4;

Qy	2	GGGGGCGCTGGGACAGATGAGCAGTGAACAGAGGCCACAGAGGCGCGTGGGGCCCTC	80
Db	17	GAGGCGCCGAGCAGAGATGAGCGGTGAATCAGAGCCACACCGAGAACCGCGGAGGCCCTTC	76
Qy	81	ATCCCTCTGCGCAAGTGTCTTGAAGCAGTGTGAGATGTGACCTCTGCTCTCTACAG	140
Db	77	ATCCCTAAGGTGAAGTCTCTTGAAGGCGTCTCGAATGTGTGAGCTCTCTTCCACAG	136
Qy	141	AAACCGTGAATCTATCTCTTTAATGCAACAAGAAAGAACGTTGTCTCACTTCA	200
Db	137	CGATCAGAAAGCTCAATGTCTTTAGTGTGAAGACAGAGAACATGTGTCTCACTTCA	196
Qy	201	TACCGGGGAGTCTTCGTGACTTCAACTAGTCAATGACCCGATGCTTCTTTATGATG	260
Db	197	TACCGGGATTTTTCATACCTTCACTGCTCCATCACTGATGCCAGTGTGTCTTTATGATG	256
Qy	261	CCGTTTGGCTGATGAGTGAAGTGCACCATTTGAACAACCAATTTTGTCCCCACACTACAT	320
Db	257	CCATTTGATGTGATGACGAGACCTGCAGTGTGAACAACAGTATTTGTGCACAACTTCAT	316
Qy	321	AAAGAACCATTTGACGACGCTCCAGGTGTGTGCTGGGAAGACAAAGCTGTTTTAAGTTA	380
Db	317	AAGGAACTATTACGAGCAGCTCCAAATGTGTGCTGGGAAGACAAAGCTTAAATTA	376
Qy	381	TCCTTTAGGAAGAGGTGCATCGAATTTGTGCCCACTGATGTTAAAGTGCCTCTGCT	440
Db	377	GTCCTTCAGAAATGAGGTGCCATTGAATTTGTCCCACTGATGTTAAAGTGCCTCTGCT	436
Qy	441	GCTGCCAGGAATTCACCTTGAAGTGAATTACTGTTCGACACTTCAGACTGTAC	500

Db 437 GCTGCCGAGAGATTTCACCTTAGAACCTTTAAATGACTGGTTCAAGCTCTATGGGAATTAT 496
 Oy 501 ATATATTCTGTCCCAAGGGCTGCAGTGTGCTCTCAAGACACTTGTCCAGCATATCCA 560
 Db 497 GTATATTCTGGGAGAGGAAT---ATGTGACCTCCACAGATGCTTGTATAG----- 545
 Oy 561 ATTGTGATCTATAGGACCCCAACCAAGATATACGTGTCCAAACAGAGGGAATATAGAACT 620
 Db 546 -TTATTGTCTATAGAGCCCACTCTGAGATATAGAGGCCCACTCCCGATACAGAGCC 604
 Oy 621 CCACCAAGAGATATAGAGCCCAACCAAGGGGAATATAGAGCCCACTTATGGATATAGA 680
 Db 605 CCACCTGCAGAGATATAGAGCCCAACCTGTAGGAATATAGAGCCCGCTGTGGATACAGA 664
 Oy 681 GCCCGGCTGTGGGATATAGAGTCCCACTAGGGGGATATAGAGTCCCACTGGGGGATAT 740
 Db 665 GCTTCACTGTGGATATAGAGCCCACTCTTGGATACGAGGCCCACTTGCAGATAT 724
 Oy 741 GAGTCTCCACTTGGGGGATATAGAGCCCACTTGGGGATATAGAGTCCCACTTGGGGGA 800
 Db 725 GAGGCCCACTTAGAGATATAGAGCCCACTTGGATATAGAACCCCACTCTTGGGA 784
 Oy 801 TATGTGTCCCACTTGGGGATATAGAGCCCACTGCAGAGATATAGAGGCCCAACCAAGT 860
 Db 785 TATGAGCCCACTCTTGGATATAGAGCCCACTGCAGAGATATAGAGGCCCGGCTGTGG 844
 Oy 861 GGAATAGAGCCCTTACCCCTGCATATAGAGCTCAATGTGCTGGAATATACAGTGTCTT 920
 Db 845 GGAATACAGAGCTTCACTGTGATCAGAGCCAGGCTTCAAGATCTACAGCAGCC--- 901
 Oy 921 CACAGATCTATGACAGTCAAGAGGAGAGCTTCTTCCCACTTACCTCATCTTCTAGGTC 980
 Db 902 -----CAGGCTCTGAAAAAGAGGCTTCTTCTCTGTGCTCTTCTTCAAGGTC 952
 Oy 981 CATTTACCACTTCTCAGAGTTAACTTTGAGATCAACCAAG--GGCACTCTAAA 1039
 Db 953 CATCTTAACCTTCTAAGATGTAACTTTAAGCTCACCAAGCAAAAGAGTACCTTAAA 1012
 Oy 1040 ACTGAATCTCAGTAAGAGGAAGCCCAAGT 1071
 Db 1013 ATTGAAGTCAAGATTAAGAGAGAGCACTCAAGT 1044

```

RESULT 3
BC022549
LOCUS
DEFINITION
BC022549 Homo sapiens, similar to RIKEN cDNA 4930511i3 gene, clone
MGC:26828 IMAGE:4815849, mRNA, complete cds.
BC022549
ACCESSION
BC022549
VERSION
BC022549.1 GI:18490710
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS
1 (bases 1 to 2266)
Strausberg, R.
TITL
Direct Submission
JOURNAL
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT
Contact: MGC help desk
Email: ggaibo-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CNSA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CNSA library Arrayed by: The I.M.A.G.E. Consortium (LUNV)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-hngc.stanford.edu

```

Contact: (Dickson, Mark) med@pax1.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnl.gov>
 Series: IRAX Plate: 32 Row: m Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES
 source
 Location/Qualifiers
 1..2266
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:26828 IMAGE:4815949"
 /issue_type="Brain, hippocampus"
 /clone_lib="NH MGC_95"
 /lab_host="DH10B"
 /note="Vector: pBluescript"

CDS

32..634
 /codon_start=1
 /product="Similar to RIKEN cDNA 4930521123 gene"
 /protein_id="AAH2549.1"
 /db_xref="GI:18490711"
 /translation="MAVNSHTENRRGALIPNGESILKRSFNVELSPFAPRBSGVNPS
 GRKTEGLFVSTRVIFPTSCSIDPMLSPMPDPDLMTNLTVBQVPAAPRFGITQIA
 PYGEGQATFLVFNQAGIAPQALMVAASAAAPLARTLDMFSSMGIVYTGK
 GICALHRCVQLSMEPHLQDMEPHLQDMEPHLQDMEPHP"

BASE COUNT 709 a 477 c 468 g 612 t
 ORIGIN

Query Match 41.6%; Score 587.2; DB 9; Length 2266;
 Best Local Similarity 75.4%; Pred. No. 1.2e-134;
 Matches 793; Conservative 0; Mismatches 233; Indels 26; Gaps 4;

```

QY 21 GGGGGCGTGGGAGATGCGAGTGAACGACGACCAACGAGAGCGCTGCTGGGCGCCCTC 80
DB 17 GAGGCGCCGAGAGAGTGGCGGTGAATCAGAGCCACGAGAAACCGCGGAGGCGCTC 76
QY 81 ATCCCTCTGGCGAAAGTGTCTTGAAGAGTGAAGATGAGACCTTGCTTCTTACAG 140
DB 77 ATCCCTAAGCGTGAAGTCTCTTGAAGCGGTCTCCGAATGTGAGGCTCTCTTCCACAG 136
QY 141 AAACGAGTGAATCTATCTCTTTAATGCGACAAAGAAAGAAAGCGTTGTTCTCACTTCA 200
DB 137 CCATCAGAGGCTCAATATGCTTAGTGTAAGAAAGACAGAAATGTTTCTCACTTCA 196
QY 201 TACCGGCTGTCTCTGTGACTTCACTTGAATGACATGACCCCATGCTTCTTTAATGAG 260
DB 197 TACCGGCTGTATTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCA 256
QY 261 CGGTTGGCTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 320
DB 257 CCATTTGATCTGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 316
QY 321 AAAGAAACCATTCAGGAGCTCCAGGTGTGCTGCGAAGCAAGCTGTTTAAAGTGA 380
DB 317 AAGGAAACCATTCAGGAGCTCCAGGTGTGCTGCGAAGCAAGCTGTTTAAAGTGA 376
QY 381 TCCCTTGAAGAAAGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 440
DB 377 GTCTTCAAGAAATGAGAGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 436
QY 441 GCGCGAGAGGAATTCAGTGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 500
DB 437 GCGCGAGAGGAATTCAGTGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 496
QY 501 ATTAATTAAGTGTGCGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 560
DB 497 GTAAATTAAGTGTGCGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 544
QY 561 ATTGTGATCTAATGAGACCCCAACGAGATATACAGTCCAAACGAGGGAATATGGAAT 620

```

```

DB 545 -TTATGTCTATGAGAGCCCACTGCAAGATATGAGAGCCCACTCCCGAATACGAGCC 603
QY 621 CCACGAGAGATATGAGAGCCCACTGCAAGATATGAGAGCCCACTATGAGATATGAG 680
DB 604 CCACCTGAGAGATATGAGAGCCCACTGCAAGATATGAGAGCCCACTGTGAGATACAG 663
QY 681 GCCCGGCTGTGAGATATGAGAGTCCCACTGAGGATATGAGAGTCCCACTGAGGATAT 740
DB 664 GCTTCACTGTGAGATATGAGAGCCCACTTGTGATGAGAGCCCACTGCAAGATAT 723
QY 741 GAGTGTCCCACTGAGGATATGAGAGCCCACTGAGGATATGAGAGTCCCACTGAGGAT 800
DB 724 GAGGCGCCCACTGTGAGATATGAGAGCCCACTTGTGATATGAGAGCCCACTTGTGAG 783
QY 801 TATGAGTCCCACTGTGAGATATGAGAGCCCACTGAGGATATGAGAGCCCACTGAGCT 860
DB 784 TATGAGAGCCCACTGTGAGATATGAGAGCCCACTGAGGATATGAGAGCCCACTGAGCT 843
QY 861 GGAATATGAGAGCCCACTGTGAGATATGAGAGCTGCTGAGAAATACAGTGTCTCT 920
DB 844 GATATCAGAGCTTCACTGTGAGATATGAGAGCCCACTGAGATATACAGAGCC--- 900
QY 921 CACAGATCTATGAGAGCTCAGAGAGAGACTTCTTCCCACTCACTCACTTCTTATGAGTC 980
DB 901 -----CAGGCTCTGAGAAAGAGAGCTTCTTCTCTCTGCTCTCTCTCTCTCTGAGTC 951
QY 981 CATTATCACTTCTCAGAGTAAACCTTGAAGACTCAACAGCAAG-GGACCCCTAA 1039
DB 952 CATTCTTAACCTTCTTAAGATGTAACCTTGAAGACTCAACAGCAAGAGGATACCTTAA 1011
QY 1040 ACTGAAGTCAAGTAAAGAGAGAGCCAGGT 1071
DB 1012 ATTGAAGTCAAGATTAAGAGAGAGAGCTCAGGT 1043

```

RESULT 4
 LOCUS AX359665 1001 bp DNA linear PAT 13-FEB-2002
 DEFINITION Sequence 11 from Patent WO0190185.
 ACCESSION AX359665
 VERSION AX359665.1 GI:18675411
 KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

1
 Oko, R. and Sutovsky, P.
 Pt32 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses thereof
 Patent: WO 0190185-A 11 29-NOV-2001;
 QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES UNIVERSITY (US)

FEATURES
 source
 Location/Qualifiers
 1..1001
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

CDS

1..708
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD23337.1"
 /db_xref="GI:18675412"
 /translation="MPPDLMNTLVBOVPANPIKGTIOAPYGMGEOATFLVPR
 NGDAIRPACLVKASAVRGRPLRTGNDWSSGIVYIRNGAKCTQMGCSYVVG
 APPRAGAPPPGTGAPPAAGAGVAGNBPVGTGAPVRRGAPPLGIGAPPAAGTAP
 PLGAGAPPLGIGTGPLGIGAPPAAGNBPVGTGAPPAAGNBPVGTGAPPAAGNBPVGTGAP
 PENRSLPSASSBOVHS"

BASE COUNT 261 a 254 c 248 g 238 t
 ORIGIN

Query Match 30.8%; Score 435.4; DB 6; Length 1001;

[illegible]

```

AUTHORS      Tracey A.
TITLE        Direct Submission
JOURNAL      Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton
              Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
              humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
              On May 26, 2003 this sequence version replaced g1:31043704.
COMMENT      ----- Genome Center
              Center: Wellcome Trust Sanger Institute
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquerry@sanger.ac.uk
              ----- Project Information
              Center project name: BT121D1
              ----- Summary Statistics
              Assembly program: XGAP4; version 4.5
              Chemistry: Dye-terminator; 100% of reads
              Consensus quality: 177502 bases at least Q40
              Consensus quality: 177971 bases at least Q30
              Consensus quality: 178283 bases at least Q20
              Insert size: 178722; sum-of-contigs
              Insert size: 167315; 12.1% error; agarose-fp
              Quality coverage: 7.82x in Q20 bases; sum-of-contigs Quality
              coverage: 8.59x in Q20 bases; agarose-fp
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 6 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1. 5850: contig of 5850 bp in length
              *
              * 5851 5950: gap of 100 bp
              *
              * 5951 74264: contig of 68314 bp in length
              *
              * 74265 74364: gap of 100 bp
              *
              * 74365 84656: contig of 10292 bp in length
              *
              * 84657 84756: gap of 100 bp
              *
              * 84757 108493: contig of 23737 bp in length
              *
              * 108494 108593: gap of 100 bp
              *
              * 108594 114454: contig of 5861 bp in length
              *
              * 114455 114554: gap of 100 bp
              *
              * 114555 179222: contig of 64668 bp in length.
              *
              Location/Qualifiers
              source
              1. 179222
              /organism="Sus scrofa"
              /mol_type="genomic DNA"
              /db_xref="taxon:9823"
              /clone="PIgR-121D21"
              /clone_1b="PIgR"
              1. 5850
              /note="assembly fragment:02838
              fragment chain:1"
              5951..74264
              /note="assembly fragment:01479
              fragment chain:1"
              74365..84656
              /note="assembly fragment:01113
              fragment chain:1"
              84757..108493
              /note="assembly fragment:00846
              fragment chain:1"
              108594..114454
              /note="assembly fragment:01058
              fragment chain:2"
              114555..179222
              /note="assembly fragment:02416
              fragment chain:2"
              44880 C 44117 G 44831 C 502 others

BASE COUNT  44892 A 44880 C 44117 G 44831 C 502 others
ORIGIN
Query Match 24.6%; Score 347.2; DB 2; Length 179222;
Best Local Similarity 76.8%; Pred No. 3,5e-75;

```

Matches 424; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 520 CTGCAGTGTCTCTCTCAAGACACCTTTCACGATATCAATTTGATCTATAGACCC 579

DB 55353 CTCACAGATACGAGGCCACCTCCAGATATCCAGACCCCACTCCAGATATCGAGCCC 55412

QY 580 CACCACAGATATACATCTCAACCGAGGGAATATGAACTCCACAGAAAGATATGAG 639

DB 55413 CACCTCAGATACGAGGCCACCTGAAAGATCTCAAGGCCCAAGAGATATGAG 55472

QY 640 CCCAACAGGAGGATATGAGCCCACTATGAGATATGAGCCCGCTGTGAGATATG 699

DB 55473 CCCATCTGCGGATATGAAAGCTCACTGCGAGATATGACGCCCTGCTCCAGATATG 55532

QY 700 GAGTCCCACTGGGAGATATGAGTCCCACTGGGAGATATGAGTCCCACTGGGAGAT 759

DB 55533 GAGCTCACCTGAAAGACAGAAAGCCCAAGAGATATGAGTCCCACTGGGAGAT 55592

QY 760 ATGAGGCCCCACCTGGGAGATATGAGTCCCACTGGGAGATATGAGTCCCACTGGGAG 819

DB 55593 ATGAGGCCCCACCTGCAAGATATGATGCCGCTCCAGATATGAGTCCCACTGGGAG 55652

QY 820 GATATGAGGCCCACTGCAAGATATGAGCCCAAGAGATATGAGTCCCACTGGGAG 879

DB 55653 GATATGAGGCCCTGCGAGATATGAGCACAAGAGTGGAAATGAGGCCCACTG 55712

QY 880 CTGCATATGAGTCCATCTGCTGGAATATGAGTCCCTTCAAGATCTATAGACGCTC 939

DB 55713 TACCATATGAAAGCCCACTGCTGGAATATGAGTCCCTTCAAGATCTATAGACGCTC 55772

QY 940 AGCAGAGATCTCTCTCCCACTACCTCATCTTATAGTCCATTTACCACTTCCAGA 999

DB 55773 AGCCGAGGCTTCTCTCTCTCACTCATCTTCTCAGGCCCATTTACCACTTCTAGA 55832

QY 1000 GTTAAACCTTGAAGACTCACCAAGCAAGGACCCCTTAAACTGAATGACAGTAAAGAG 1059

DB 55833 AGTAACTTGAAGATTTCAACCAAGCAAGGATCCCTTAAACTGAATGACAGTAAAGAG 55892

QY 1060 GAAGACCCAGGT 1071

DB 55893 GAGGACTCAGGT 55904

RESULT 6
BX470149 222469 bp DNA linear HTG 07-MAY-2003
LOCUS
DEFINITION Danio rerio clone DKX-242K7, *** SEQUENCING IN PROGRESS ***, 50
unoriented pieces.
ACCESSION BX470149
VERSION BX470149.3 GI:30424228
KEYWORDS HTG; HTGS PHASBL.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Bkayote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 222469)
Burton, J.
Direct Submission
Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On May 7, 2003 this sequence version replaced gi:30387077.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zK242K7
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Chemistry: Dye-terminator; 100% of reads

Consensus quality: 197958 bases at least Q40
Consensus quality: 206715 bases at least Q30
Consensus quality: 212351 bases at least Q20
Insert size: 217569; sum-of-contigs
Insert size: 165837; 6.2% error; agarose-fp
Quality coverage: 2.48x in Q20 bases; sum-of-contigs quality
coverage: 3.95x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6657 6656: contig of 6656 bp in length
6757 6756: gap of 100 bp
10255 10255: contig of 3499 bp in length
10256 10355: gap of 100 bp
10356 13323: contig of 2868 bp in length
13224 13323: gap of 100 bp
13324 13935: contig of 6612 bp in length
13936 20035: gap of 100 bp
20036 22597: contig of 2562 bp in length
22598 22697: gap of 100 bp
22698 28569: contig of 5872 bp in length
28570 28699: gap of 100 bp
28670 37965: contig of 9236 bp in length
37966 38065: gap of 100 bp
38066 41140: contig of 3075 bp in length
41141 41240: gap of 100 bp
41241 45108: contig of 3868 bp in length
45109 45208: gap of 100 bp
45209 47730: contig of 2522 bp in length
47731 47830: gap of 100 bp
47831 51884: contig of 4054 bp in length
51885 51984: gap of 100 bp
51985 55957: contig of 3873 bp in length
55958 55957: gap of 100 bp
55959 60498: contig of 4541 bp in length
60499 60598: gap of 100 bp
60599 63641: contig of 3043 bp in length
63642 63741: gap of 100 bp
63742 72674: contig of 8933 bp in length
72675 72774: gap of 100 bp
72775 75602: contig of 2828 bp in length
75603 75702: gap of 100 bp
75703 78062: contig of 2360 bp in length
78063 78162: gap of 100 bp
78163 84485: contig of 6323 bp in length
84486 84585: gap of 100 bp
84586 87553: contig of 2968 bp in length
87554 87653: gap of 100 bp
87654 92487: contig of 4834 bp in length
92488 92587: gap of 100 bp
92588 96104: contig of 3517 bp in length
96105 96204: gap of 100 bp
96205 101479: contig of 5275 bp in length
101480 101579: gap of 100 bp
101580 104979: contig of 3400 bp in length
104980 105079: gap of 100 bp
105080 108003: contig of 2924 bp in length
108004 108103: gap of 100 bp
108104 113151: contig of 5048 bp in length
113152 113251: gap of 100 bp
113252 116585: contig of 3334 bp in length
116586 116685: gap of 100 bp
116687 119346: contig of 2661 bp in length
119347 119446: gap of 100 bp
119447 123125: contig of 3679 bp in length
123126 123225: gap of 100 bp
123226 131266: contig of 8041 bp in length

[illegible]

```
source
1..222469
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKRY-242K"
/clone_1b="DarioKey"
1..6656
/note="assembly_fragment:01359
fragment_chain:1"
6757..10255
/note="assembly_fragment:01385
fragment_chain:1"
10356..13223
/note="assembly_fragment:00592
fragment_chain:1"
13324..19935
/note="assembly_fragment:00237
fragment_chain:1"
20036..22597
/note="assembly_fragment:00949
fragment_chain:1"
22698..28569
/note="assembly_fragment:00615
fragment_chain:1"
28670..37965
/note="assembly_fragment:01306
fragment_chain:2"
38066..41140
/note="assembly_fragment:00487
fragment_chain:2"
```

```
misc_feature      41241.45108
/notes="assembly_fragment:00232
fragment_chain:2"
45209.47730
/notes="assembly_fragment:00935
fragment_chain:2"
47831.51884
/notes="assembly_fragment:01415
fragment_chain:2"
51985.55857
/notes="assembly_fragment:00878
fragment_chain:3"
55958.60498
/notes="assembly_fragment:01208
fragment_chain:3"
60599.63641
/notes="assembly_fragment:01266
fragment_chain:3"
63742.72674
/notes="assembly_fragment:01329
fragment_chain:3"
72775.75602
/notes="assembly_fragment:01183
fragment_chain:4"
75703.78062
/notes="assembly_fragment:01179
```

Query Match	24.6%	Score 347.2;	DB 2;	Length 222469;
Best Local Similarity	76.8%	Pred. No. 3.5e-75;		
Matches 424;	Conservative 0;	Mismatches 128;	Indels 0;	Gaps 0;

QY	520	TTGAGTGTGTGCTCCCTCACAAGACCTTGCTCCAGCATATCAATTGATGATCTATGAGCCC	579
Db	197009	CTCCAGGATATAGGAGGCCCCCACTTCAGAGATACGAGGCCCCCACTTCAGAGATACGAGCCC	197068
QY	580	CACCAACCGAGATATACAGTCCACCAAGGGAATATGAACTCCACCAAGAGATATGAG	639
Db	197069	CACCTCCAGGATACGAGGCCCCCACTGAAAGACTGGAAGCCCCCAACAGCAGATATGAG	197128
QY	640	CCCAACGAGGGAGATATGAGGCCCCCACTATGGAATATGGAAGCCCGGCTGTGGGATATG	699
Db	197129	CCCCATCTGCGGGATATGAGCTCACTTGAGAGATATGACGCCCTTGCTTCAGAGATATG	197188
QY	700	GAGTCCCACTGGGGGATATGAGTCCCACTGGGGGATATGAGTCCCACTGGGGGAT	759
Db	197189	GAGCTCACTGAAAGACAGAGGCCCAACAGCAGAGATATGAGGTCCCATCTGCGGGAT	197248
QY	760	ATGAGAGCCCACTGGGGGATATGAGTCCCACTGGGGGATATGAGTCCCACTGGGG	819
Db	197249	ATGAGGCCCACTGCAAGATATGATGCCCCGCTTCAGAGATATGAGGCTCACTGAA	197308
QY	820	GATATGAGAGCCCACTGCGAGATATGAGGCCCACTGAGATATGAAAGCCCTACCC	879
Db	197309	GATATGAGAGCCCCCTCTGCAAGATATGAGCAACAACAGCTGGAATGAAAGCCCACTTC	197368
QY	880	CTGCATATGAAAGCTTCATCTGCTGGAATATCAGTGCCTCTCAAGATCTATGACAGCTC	939
Db	197369	TAGCATATGAAAGCCCACTGCTGGAATATGAGCTGCCTCTCACAATCTGTGGCAGCCC	197428
QY	940	AGCAGAGAGACTTCTCTCCCACTACCTCATCTTTAAGTCCATTTACCACTTCACAG	999
Db	197429	AGCCGAGAGCTTCTCTCTCTTACTCATCTTCTCAGGCCCAATTTACCACTTCTAAG	197488
QY	1000	GTTAAACCTTGAAGCTCAACCAAGCAAGGCACTCTAAATCTGAAGTCAAGTAAAG	1059
Db	197489	AGTTAAACCTTAAAGTATCAACAAGCAAGGATACCTTAAATCTGAAGTCAAGTAAAG	197548
QY	1060	GAAGAGCCAGGT 1071	
Db	197549	GAGAGCTAGGT 197560	

HS250D10 220895 bp DNA linear PRI 05-JUN-2003
 LOCUS
 DEFINITION Human DNA sequence from clone CTA-250D10 on chromosome 22 contains the genes for SREBF2 (sterol regulatory element binding transcription factor 2), NGA (alpha-N-acetylgalactosaminidase), a gene similar to neuronal-specific septin 3, a pseudogene similar to ANT2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a genomic marker D2S1178, a CA repeat polymorphism, ESTs and a Cpg island, complete sequence.
 Z99716
 VERSION Z99716.4 GI:4456457
 KEYWORDS HTG; ANT2; Cpg Island; D2S1178; NGA; septin 3; SREBF2.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 220895)
 REFERENCE
 AUTHORS Clark, G.
 TITLE Direct Submission
 JOURNAL Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Mar 21, 1999 this sequence version replaced GI:4164339.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep/CTA-250D10 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.
 VECTOR: pBAC1081
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr22
 This sequence is the entire insert of clone CTA-250D10. The true left end of clone RPI-18601 is at 129979 in this sequence. The true left end of clone RPI-359016 is at 1335 in this sequence. The true right end of clone RPI-821D11 is at 23458 in this sequence. The true right end of clone RPI-359016 is at 118711 in this sequence.
 Location/Qualifiers
 1..220895
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="CTA-250D10"
 /clone_1lb="CTG78SK-A1"
 17..29
 /note="2.6 copies 5 mer ATTGG 26% conserved"
 repeat_region
 33..325
 /note="AluX repeat: matches 1..301 of consensus"
 repeat_region
 326..338

repeat_region
 667..683
 /note="2.1 copies 8 mer AGAAGACC 34% conserved"
 repeat_region
 740..751
 /note="2.4 copies 5 mer TTTTA 24% conserved"
 repeat_region
 1443..1457
 /note="7.5 copies 2 mer AC 21% conserved"
 repeat_region
 1715..1997
 /note="Alu repeat: matches 5..293 of consensus"
 repeat_region
 2103..2378
 /note="AluX repeat: matches 2..296 of consensus"
 repeat_region
 2844..2855
 /note="12.0 copies 1 mer A 24% conserved"
 repeat_region
 complement(2884..3033)
 /note="MIR repeat: matches 70..235 of consensus"
 repeat_region
 complement(3040..3344)
 /note="Alu repeat: matches 1..310 of consensus"
 repeat_region
 3446..3681
 /note="AluX repeat: matches 5256..5471 of consensus"
 repeat_region
 3746..3837
 /note="AluX repeat: matches 5615..5707 of consensus"
 repeat_region
 3838..4142
 /note="Alu repeat: matches 2..305 of consensus"
 repeat_region
 4143..4441
 /note="AluX repeat: matches 5707..5999 of consensus"
 repeat_region
 4442..4747
 /note="AluX repeat: matches 1..292 of consensus"
 repeat_region
 4748..4858
 /note="AluX repeat: matches 5999..6110 of consensus"
 repeat_region
 4881..5172
 /note="AluX repeat: matches 1..296 of consensus"
 repeat_region
 5418..5429
 /note="2.4 copies 5 mer GCTGA 24% conserved"
 repeat_region
 complement(5769..6021)
 /note="match: GGS: km:A0069839"
 repeat_region
 6074..6089
 /note="2.0 copies 8 mer CCCGCTG 32% conserved"
 repeat_region
 6139..6153
 /note="2.5 copies 6 mer TTTTA 21% conserved"
 repeat_region
 6141..6155
 /note="3.0 copies 5 mer TTTTA 21% conserved"
 repeat_region
 complement(6366..6572)
 /note="MIR repeat: matches 5..214 of consensus"
 repeat_region
 6839..6852
 /note="2.3 copies 6 mer GAATT 28% conserved"
 repeat_region
 7088..7114
 /note="6.8 copies 4 mer TTTA 36% conserved"
 repeat_region
 7091..7109
 /note="1.9 copies 10 mer ATTATTTAC 38% conserved"
 repeat_region
 complement(7122..7418)
 /note="AluX repeat: matches 1..300 of consensus"
 repeat_region
 7464..7502
 /note="19.5 copies 2 mer AT 78% conserved"
 repeat_region
 7502..7524
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region
 7523..7538
 /note="2.0 copies 8 mer CAAATAT 32% conserved"
 repeat_region
 7535..7563
 /note="14.5 copies 2 mer AT 49% conserved"
 repeat_region
 complement(7564..7859)
 /note="Alu repeat: matches 1..296 of consensus"
 repeat_region
 7873..7893
 /note="5.2 copies 4 mer TGT 33% conserved"
 repeat_region
 7914..8200
 /note="AluX repeat: matches 1..290 of consensus"
 repeat_region
 complement(8202..8318)
 /note="AluX/FLM repeat: matches 1..117 of consensus"
 repeat_region
 8387..8409
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region
 8483..8498
 /note="2.0 copies 8 mer GTGTGT 32% conserved"
 repeat_region
 8486..8496
 /note="2.2 copies 5 mer GTGT 22% conserved"

```

repeat_region      8606..8622
                    /note="2.4 copies 7 mer AGAGAAA 34% conserved"
repeat_region      complement(8767..8899)
                    /note="MER115 repeat: matches 33..268 of consensus"
repeat_region      9004..9014
                    /note="2.2 copies 5 mer CTGCG 22% conserved"
repeat_region      9037..9140
                    /note="WIR repeat: matches 17..123 of consensus"
repeat_region      9509..9823
                    /note="AluXk repeat: matches 1..312 of consensus"
repeat_region      9872..9883
                    /note="2.4 copies 5 mer GTCTT 24% conserved"
gene               10071..49606
                    /gene="SRBP2"
mRNA               join(<10071..10520,11851..12032,14129..14275,17038..17259,
                    18577..18691,18783..18964,20469..20661,21182..21363,
                    23956..24232,28082..28251,36357..36528,38060..38177,
                    40292..40401,41889..42021,43570..43738,46210..46395,
                    48103..48214,48680..48900)
                    /gene="SRBP2"
                    /product="bK250D10.1 (sterol regulatory element binding
                    /transcription factor 2)"
                    /note="sterol regulatory element binding transcription
                    factor 2"
match: cDNAs: Bm:U02031 Bm:U12330 Bm:U00968 Bm:U09103
Bm:U16995 Bm:U22818 Bm:U22819 Bm:U12329 Bm:AB017337
match: ESTs: Bm:R36312 Bm:AL040138 Bm:H92808 Bm:AA053886"
/evidence=not_experimental
join(<10071..10520,11851..12032,14129..14275,17038..17259,
18577..18691,18783..18964,20469..20661,21182..21363,
23956..24232,28082..28251,36357..36528,38060..38177,
40292..40401,41889..42021,43570..43738,46210..46395,
48103..48214,48680..48900)
/gene="SRBP2"
/note=match: proteins: Sw:Q12772 Sw:Q60429 Sw:P36956
Sw:Q60416 Tr:O15352"
/codon_start=3
/evidence=not_experimental
/product="bK250D10.1 (sterol regulatory element binding
transcription factor 2)"
/protein_id="CAB41234.1"
/db_xref="gi:5578715"
/translaction="MLOQVSNVGRBPRDLPBRCQSSPRGSGSSGSSGSSSS
NGRSSGAVDPVSORSTQVTLPSFSSAASPOAPTLQVVSPTSPTPRATPILQ
PRPQPPOPTOLOQOQVMTITPTSTVQTRITQPLTYQAATSFOVLQVQSLT
SSOVQPTVIOOVQVQVQAVLTQVANGTQLAPATVQVVAQVQVQVQVQVQVQV
KTDLVLTTLTKDSPVAAVONPALTLTPTQALQVQVQVQVQVQVQVQVQVQV
GQKVPVKOVGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
HKSGLAKAIDIVITQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
DFQNVLMSPSPADSGSQAGPSPSIDSEPSFLDDAKYKDEPDEFPVALGAVDSS
RILCVLFLCLSPNPLSLQWQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
LILMVGAVVILSVFVKLVHGEVIRPHSRSSVTFMHRRQADLDLARGPAAAGN
LQTCGAVGALPTSRDLQSLSNVIRYSIQRLVRYMLKVKVQVQVQVQVQVQVQV
PDRKATSRAPALAAVYRHQIHQITGCLPAGSACDYMALCAVNLARCAKRIKPTP
LVBIHTRAMGLKTRCGKGLFSLASPIFSRAQSLCGPSPHSAVPDSIRKLCPLOKPP

```

Query Match 20.0% Score 282.6; DB 9; Length 220895;
 Best Local Similarity 74.6% Pred. No. 3.7e-59;
 Matches 387; Conservative 0; Mismatches 119; Indels 13; Gaps 2;

```

QY 554 ATATCCATTGTGATCTATGACCCCAACACAGATATATACCTCAACAGAGGAGATA 613
DB 169998 ATTCCTCATTTGTCTATGAGCCCACTGACAGATATGAGGCCCACTCCCGAGTA 170057
QY 614 TGAAGCTCCAGAGATATGAGCCCAACAGAGGAGATATGAGGCCCACTCACTGAG 673
DB 170058 CGAGACCCCACTGAGATATGAGCCCAACAGAGGAGATATGAGGCCCACTCTGAG 170117
QY 674 ATATGAGACCCCACTGAGATATGAGCTCCCACTGAGGAGATATGAGCTCCCACTGAG 733
DB 170118 ATACAGACCTCACTGAGATATGAGGCCCACTCTGAGATATGAGGCCCACTGAG 170177
QY 734 GAGATATGAGATCCCACTGAGGAGATATGAGGCCCACTCTGAGGAGATATGAGCTCCCACT 793

```

```

DB 170178 AGGATATGAGACCCCACTCTGAGATATGAGGCCCACTCTGAGATATGAGGCCCACT 170237
QY 794 TGGGAGATATGAGTCCCACTCTGAGGAGATATGAGGCCCACTCTGAGATATGAGGCCCACT 853
DB 170238 TCTCGAGATATGAGGCCCACTCTGAGATATGAGGCCCACTCTGAGGAGATATGAGGCCCACT 170297
QY 854 ACCAGCTGAGAAATGAGGCCCACTCTGAGATATGAGGCCCACTCTGAGGAGATATGAGGCCCACT 913
DB 170298 GCTTCGAGATATGAGGCCCACTCTGAGATATGAGGCCCACTCTGAGGAGATATGAGGCCCACT 170357
QY 914 TGCCTCTCAGAGATATGAGCACTCTGAGGAGATATGAGGCCCACTCTGAGGAGATATGAGGCCCACT 973
DB 170358 AGCC-----CAGGCTCTGAGAAACAGAGGCTCTCTTCCCTCTGAGGCTCTCTT 170405
QY 974 TTAGGCTCATTTACCACTCTCTGAGGATTAACCTTGAAGATCAACAGAGGAG-AGCA 1032
DB 170406 TCAAGCTCATTTCTTAACTCTTAAAGATTAACCTTGAAGATCAACAGAGGAGGTA 170465
QY 1033 CCTTAAATGAGATCACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1071
DB 170466 CCTTAAATGAGATCACTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 170504

```

RESULT 8

LOCUS BX296540 129624 bp DNA linear HTG 02-APR-2003
 DEFINITION unsorted pieces.
 ACCESSION BX296540
 VERSION BX296540.2 GI:29500961
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 1 (bases 1 to 129624)

REFERENCE
 AUTHORS Burton, J.
 TITLE Direct Submission
 JOURNAL Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Apr 2, 2003 this sequence version replaced gi:29335441.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: b7231K18
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 113740 bases at least Q40
 Consensus quality: 118187 bases at least Q30
 Consensus quality: 120712 bases at least Q20
 Insert size: 154528; 4.6% error; agarose-ff
 Quality coverage: 2.35x in Q20 bases; sum-of-contigs Quality
 coverage: 3.19x in Q20 bases; agarose-ff

NOTE: This is a 'working draft' sequence. It currently
 consists of 37 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 * 1 4262: contig of 4262 bp in length
 * 4263 4362: gap of 100 bp
 * 4363 6991: contig of 2629 bp in length
 * 6992 7091: gap of 100 bp

FEATURES	SOURCE
7092	9480: contig of 2389 bp in length
9481	9580: gap of 100 bp
9581	11688: contig of 2108 bp in length
11688	11788: gap of 100 bp
11789	14658: contig of 2870 bp in length
14659	14758: gap of 100 bp
14759	18440: contig of 3682 bp in length
18441	18540: gap of 100 bp
18541	20818: contig of 2278 bp in length
20819	20918: gap of 100 bp
20919	22969: contig of 2051 bp in length
22970	23069: gap of 100 bp
23070	27584: contig of 4515 bp in length
27585	27684: gap of 100 bp
27685	33211: contig of 5527 bp in length
33212	33311: gap of 100 bp
33312	35521: contig of 2210 bp in length
35522	35621: gap of 100 bp
35622	38528: contig of 2907 bp in length
38529	38628: gap of 100 bp
38629	42063: contig of 3435 bp in length
42064	42163: gap of 100 bp
42164	44164: contig of 2001 bp in length
44165	44264: gap of 100 bp
44265	47707: contig of 3443 bp in length
47708	47807: gap of 100 bp
47808	50737: contig of 2930 bp in length
50738	50837: gap of 100 bp
50838	57066: contig of 6229 bp in length
57067	57166: gap of 100 bp
57167	59978: contig of 2812 bp in length
59979	60078: gap of 100 bp
60079	62560: contig of 2482 bp in length
62561	62660: gap of 100 bp
62661	64963: contig of 2303 bp in length
64964	65063: gap of 100 bp
65064	67191: contig of 2128 bp in length
67192	67291: gap of 100 bp
67292	72005: contig of 4714 bp in length
72006	72105: gap of 100 bp
72106	76595: contig of 4490 bp in length
76596	76695: gap of 100 bp
76696	79825: contig of 1130 bp in length
79826	79925: gap of 100 bp
79926	83540: contig of 3615 bp in length
83541	83640: gap of 100 bp
83641	86276: contig of 2636 bp in length
86277	86376: gap of 100 bp
86377	94620: contig of 8244 bp in length
94621	94720: gap of 100 bp
94721	97520: contig of 2800 bp in length
97521	97620: gap of 100 bp
97621	99784: contig of 2164 bp in length
99785	99884: gap of 100 bp
99885	102112: contig of 2228 bp in length
102113	102212: gap of 100 bp
102213	104242: contig of 2030 bp in length
104243	104342: gap of 100 bp
104343	110710: contig of 6368 bp in length
110711	110810: gap of 100 bp
110811	113367: contig of 2557 bp in length
113368	113467: gap of 100 bp
113468	115505: contig of 2038 bp in length
115506	115605: gap of 100 bp
115606	121874: contig of 6269 bp in length
121875	121974: gap of 100 bp
121975	127096: contig of 5122 bp in length
127097	127196: gap of 100 bp
127197	129624: contig of 2428 bp in length.
Location/Qualifiers	
1..129624	
/organism="Sus scrofa"	
/mol_type="genomic DNA"	

FEATURES	SOURCE
misc_feature	/db xref="taxon:9823"
misc_feature	/clone="P1G8-231X18"
misc_feature	/clone_11b="P1G8"
misc_feature	1..4262
misc_feature	/note="assembly_fragment:00361"
misc_feature	fragment chain:1"
misc_feature	4363..6591
misc_feature	/note="assembly_fragment:00108"
misc_feature	fragment chain:1"
misc_feature	7092..9480
misc_feature	/note="assembly_fragment:00460"
misc_feature	fragment chain:1"
misc_feature	9581..11688
misc_feature	/note="assembly_fragment:00959"
misc_feature	fragment chain:1"
misc_feature	11789..14658
misc_feature	/note="assembly_fragment:00995"
misc_feature	fragment chain:2"
misc_feature	14759..18440
misc_feature	/note="assembly_fragment:00608"
misc_feature	fragment chain:2"
misc_feature	18541..20818
misc_feature	/note="assembly_fragment:01101"
misc_feature	fragment chain:2"
misc_feature	20919..22969
misc_feature	/note="assembly_fragment:00996"
misc_feature	fragment chain:2"
misc_feature	23070..27584
misc_feature	/note="assembly_fragment:00217"
misc_feature	fragment chain:3"
misc_feature	27685..33211
misc_feature	/note="assembly_fragment:00019"
misc_feature	fragment chain:3"
misc_feature	33312..35521
misc_feature	/note="assembly_fragment:00548"
misc_feature	fragment chain:3"
misc_feature	35622..38528
misc_feature	/note="assembly_fragment:00815"
misc_feature	fragment chain:4"
misc_feature	38629..42063
misc_feature	/note="assembly_fragment:00246"
misc_feature	fragment chain:4"
misc_feature	42164..44164
misc_feature	/note="assembly_fragment:00859"
misc_feature	fragment chain:5"
misc_feature	44265..47707
misc_feature	/note="assembly_fragment:00964"
misc_feature	fragment chain:5"
misc_feature	47808..50737
misc_feature	/note="assembly_fragment:01192"
misc_feature	fragment chain:6"
misc_feature	50838..57066
misc_feature	/note="assembly_fragment:00542"
misc_feature	fragment chain:6"
misc_feature	57167..59978
misc_feature	/note="assembly_fragment:01304"
misc_feature	fragment chain:7"
misc_feature	60079..62560
misc_feature	/note="assembly_fragment:00948"
misc_feature	fragment chain:7"
misc_feature	62661..64963
misc_feature	/note="assembly_fragment:00016"
misc_feature	65064..67191
misc_feature	/note="assembly_fragment:00065"
misc_feature	67292..72005
misc_feature	/note="assembly_fragment:00070"
misc_feature	72106..76595
misc_feature	/note="assembly_fragment:00136"
misc_feature	76696..79825
misc_feature	/note="assembly_fragment:00162"
misc_feature	79926..83540
misc_feature	/note="assembly_fragment:00367"
misc_feature	83641..86276

/note="assembly_fragment:00492"
 misc_feature 66377..94620
 /note="assembly_fragment:00545"
 misc_feature 94721..97520
 /note="assembly_fragment:00745"
 misc_feature 97621..99784
 /note="assembly_fragment:00801"

Query Match 15.8%; Score 223.2; DB 2; Length 129624;
 Best Local Similarity 74.2%; Pred. No. 2e-44;
 Matches 282; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 550 CAGCATATCCATGTGTATGATGACCCCAACGAGATATACAGTCCAAACGAGGG 609
 DB 65511 CATCATGCCCACTGTGTGTATGACCCCTACCTGAGATGAGAGCCCAACGAG 65452
 QY 610 AATATGAACTCCACCAAGATATGAGCCCAACGAGGAGATATGAGCCCACTA 669
 DB 65451 AATATGAGCCCACTCCAGATATGAGCCCACTCCAGATATGAGCCCACTC 65392
 QY 670 TGGATATGAGAGCCCGCTGTGGATATGAGTCCCACTGGGAGATATGAGTCCAC 729
 DB 65391 CAGGATACGAGAGCCCACTCCAGATATGAGAGCCCACTCCAGATATGAGAGCCCACT 65332
 QY 730 CTGGGGGATATGAGATCCCACTGGGAGATATGAGAGCCCACTGGGAGATATGAGATCC 789
 DB 65331 CTGAAAGACTGAAACCCCAACGAGATATGAGAGCCCACTGGGATATGAGAGCTT 65272
 QY 790 CACCTGGGGGATATGATGATCCCACTGGGAGATATGAGAGCCCACTGAGATATGAGAG 849
 DB 65271 CACCTGGGAGATATGATGATCCCTCCAGATATGAGAGCCCACTGAAAGACGAGAG 65212
 QY 850 CCCCACGAGTGAATGAAGCCCTCACTGATATGAGATGATGATGATGATGATGATGAT 909
 DB 65211 CCCCACGAGGATATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 65152
 QY 910 CAGCTGCTCTTCACGATCT 929
 DB 65151 ATGCCCGCTTCAGATAT 65132

RESULT 9
 LOCUS AC113593 168425 bp DNA linear HTG 06-JUN-2002
 DEFINITION Mus musculus clone RP23-363124, WORKING DRAFT SEQUENCE, 18 ordered
 pieces.
 AC113593
 AC113593.3 GI:21327432
 HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 168425)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP23-363124
 Unpublished
 2 (bases 1 to 168425)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
 Anderson, S., Barna, N., Baetien, V., Boguslavsky, L., Boukhvalter, B.,
 Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Dearlano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Glende, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karata, A., Kelle, C., Larocque, K., Lamazares, R.,
 Lander, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
 Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Milhova, T.,
 Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunhang, P., Pierre, N., Pollara, V., Raymond, C.,

TITLE
 JOURNAL
 REFERENCES
 AUTHORS
 Submitted (03-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 168425)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
 Anderson, S., Barna, N., Baetien, V., Bloom, T., Boguslavsky, L.,
 Boukhvalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
 Cook, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Glende, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karata, A., Kelle, C., Larocque, K., Lamazares, R.,
 Lander, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
 Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Milhova, T.,
 Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 COMMENT
 Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 6, 2002 this sequence version replaced gi:21313869.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L23816
 Center clone name: 363.1.24
 Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 162227 bases at least Q40
 Consensus quality: 164865 bases at least Q30
 Consensus quality: 165972 bases at least Q20
 Insert size: 16200; agarose-IP
 Insert size: 166725; sum-of-contigs
 Quality coverage: 8.6 in Q20 bases; agarose-IP
 Quality coverage: 8.3 in Q20 bases; sum-of-contigs
 NOTES: This is a 'working draft' sequence. It currently
 consists of 18 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces
 is believed to be correct as given, however the sizes
 of the gaps between them are based on estimates that have
 been provided by the submitter.
 This sequence will be replaced
 by the finished sequence as soon as it is available and
 the accession number will be preserved.
 1 1412: contig of 1412 bp in length
 * 1413 1512: gap of 100 bp
 * 1513 3366: contig of 1854 bp in length
 * 3367 3466: gap of 100 bp

*	3467	5542: contig of 2076 bp in length
*	5543	5642: gap of 100 bp
*	5643	7949: contig of 2307 bp in length
*	7950	8049: gap of 100 bp
*	8050	42610: contig of 34561 bp in length
*	42611	42710: gap of 100 bp
*	42711	45717: contig of 3007 bp in length
*	45718	45817: gap of 100 bp
*	45818	50106: contig of 4289 bp in length
*	50107	50206: gap of 100 bp
*	50207	52945: contig of 2739 bp in length
*	52946	53045: gap of 100 bp
*	53046	56269: contig of 3224 bp in length
*	56270	56369: gap of 100 bp
*	56370	59897: contig of 3528 bp in length
*	59898	59997: gap of 100 bp
*	59998	67207: contig of 7210 bp in length
*	67208	67307: gap of 100 bp
*	67308	78743: contig of 11336 bp in length
*	78744	78843: gap of 100 bp
*	78844	89391: contig of 11148 bp in length
*	89392	90091: gap of 100 bp
*	90092	107818: contig of 17827 bp in length
*	107919	108018: gap of 100 bp
*	108019	124874: contig of 16856 bp in length
*	124875	124974: gap of 100 bp
*	124975	142093: contig of 17119 bp in length
*	142094	142193: gap of 100 bp
*	142194	167111: contig of 24918 bp in length
*	167112	167211: gap of 100 bp
*	167212	168425: contig of 1214 bp in length

```
misc_feature /organism="Mus musculus"
misc_feature /mol_type="genomic DNA"
misc_feature /db_xref="taxon:10090"
misc_feature /clone="RP23-363I24"
misc_feature /clone_1b="RPCI-23 Female Mouse BAC"
misc_feature 1..1412
misc_feature /note="assembly_fragment"
misc_feature 1513..3366
misc_feature /note="assembly_fragment"
misc_feature 3467..5542
misc_feature /note="assembly_fragment"
misc_feature 5643..7949
misc_feature /note="assembly_fragment"
misc_feature 8050..142610
misc_feature /note="assembly_fragment"
misc_feature 42711..45717
misc_feature /note="assembly_fragment"
misc_feature 45818..50106
misc_feature /note="assembly_fragment"
misc_feature 50207..52945
misc_feature /note="assembly_fragment"
misc_feature 53046..56269
misc_feature /note="assembly_fragment"
misc_feature 56370..53897
misc_feature /note="assembly_fragment"
misc_feature 59998..67207
misc_feature /note="assembly_fragment"
misc_feature 67308..78743
misc_feature /note="assembly_fragment"
misc_feature 78844..89391
misc_feature /note="assembly_fragment"
misc_feature 90092..107918
misc_feature /note="assembly_fragment"
misc_feature 108019..124874
misc_feature /note="assembly_fragment"
misc_feature 124975..142093
misc_feature /note="assembly_fragment"
misc_feature 142194..167111
misc_feature /note="assembly_fragment"
misc_feature 167112..168425
```

Query Match	13.4%	Score 188.8	DB 2	Length 168425
Best Local Similarity	65.2%	Pred. No. 6.9e-36		
Matches 313	Conservative 0	Mismatches 157	Indels 10	Gaps 2
BASE COUNT	44391 a	39572 c	38747 g	44015 t 1700 others
ORIGIN	/note="assembly_fragment clone_end:t7 vector_side:right"			
602	ACGAGGGGATATATGAACTTCACACAGAGATATGAGAGCCCAAGAGGGGATATGAGC	661		
66302	AGCGTGAATATATGCGCTCCCTCTCTCTATATATGATACCCCAATGGGCTATGAGT	66361		
662	CCCACTTATGGATATGAGAGCCCGCTGTGGATATGAGATTCACCTGTGGGGATATGG	721		
66362	TCAACCTCCGATATATGAGCTTCACCTGTGAGATACGATCCCACTTCGATATGA	66421		
722	AGTCCCACTGGGGGATATGAGATTCACCTGGGGGATATGAGAGCCCACTGGGGGAT	781		
66422	AGCCCACTATATGAGATATGAGAGCCCACTCCAGATACGAACTACCTATGAGATC	66481		
782	TGAGATTCACCTGTGGGGATATGATGTCGCCCACTGTGGGGATATGAGAGCCCACTG	841		
66482	CGGTTCGCCACCTTCAGATATGAGAGCCCGGCTATGGGGATATGAGAACTCTCTTCA	901		
842	ATATGAGAGCCCACTGAGATATGAGAGCCCACTCCCTGATATGAGATTCATTCG	901		
66542	GCGTGAATCCATCTCTCTGATATGAGAGCCCACTCTGATATGAGAGAGCTTCACCT	9601		
902	TGGAATATGAGAGCTCTTCACAGATCTATGAGAGCTGACA-----GAGACTTC	952		
66602	TGGGTCTGAAGAGAGCCATCCAGTCTGTGAGAGTCCAGAAATCTGAATTCAGAGCAT	66661		
953	TCTTCCCACTACCTCATCTCTTATGATTCATTTACCACTTCCTGAGATTAACCTTGA	1012		
66662	TTTTCCCTCATCTCTCTCTCTCAAGATTCACCTCTCCCGCTGATATTAACCTTGA	66721		
1013	GACTACCAAGC-AAAGGGCAGCTTAAACTGAATGACATGATGAGAGAGAGCCAGGT	1071		
66722	GTTCACCAAGCAAAACAGCAGCCCTATGTGATGATGATGATGATGATGATGATGAT	66781		
RESULT 10				
AC104325	224086 bp	DNA	linear	ROD 30-MAY-2003
LOCUS	Mus musculus clone rp23-204m3 map 15 strain C57BL/6J, complete			
DEFINITION	sequence.			
ACCESSION	AC104325			
VERSION	AC104325.28	GI:31193955		
KEYWORDS	HTG.			
SOURCE	Mus musculus (house mouse)			
ORGANISM	Mus musculus			
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE	1 (bases 1 to 224086)			
JOURNAL	Jiang, H., Song, J., and Roe, B. A.			
REFERENCE	Mus musculus BAC Clone rp23-204m3			
AUTHORS	Unpublished			
TITLE	2 (bases 1 to 224086)			
JOURNAL	Jiang, H., Song, J., and Roe, B. A.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (10-DEC-2001) Department of Chemistry And Biochemistry,			
TITLE	The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,			
JOURNAL	OK 73019, USA			
REFERENCE	3 (bases 1 to 224086)			
AUTHORS	Jiang, H., Song, J., and Roe, B. A.			
TITLE	Direct Submission			
JOURNAL	Submitted (19-MAR-2003) Department Of Chemistry And Biochemistry,			
REFERENCE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,			
AUTHORS	OK 73019, USA			
TITLE	4 (bases 1 to 224086)			
JOURNAL	Jiang, H., Song, J., and Roe, B. A.			

TITLE Direct Submission
JOURNAL Submitted (14-MAY-2003) Department of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
REFERENCE 5 (bases 1 to 224086)
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Direct Submission
 Submitted (30-MAY-2003) Department of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA
COMMENT On May 30, 2003 this sequence version replaced gi:29124187.
 ----- Genome Center
 The University of Oklahoma
 Center code:UOKNOR

FEATURES
 source
 1. 224086
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /map="15"
 /clone="T23-204m3"
 /clone_1b="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"
 BASE COUNT 55592 a 55124 c 56095 g 57275 t
 ORIGIN

Query Match 13.4%; Score 188.8; DB 10; Length 224086;
 Best Local Similarity 65.2%; Pred. No. 6.8e-36;
 Matches 313; Conservative 0; Mismatches 157; Indels 10; Gaps 2;

602 ACCAGGGGATATGGAATCTCCACGAGAGATATGAGCCCAACGAGGGGATATGAGC 661
 162848 AGCTGTGAGATATGCTCCCTCCCTCTCTATATGATGATACCCCAATGGGCTATGAGT 162907
 662 CCCACTATGATATGAGGCCCCGCTGTGAGATATGAGTCCCACTGGGGATATG 721
 162908 TCCACCTCTGATATGAGCTCCACCTGTGATGATGAGATCCCACTCTCTGATATGA 162967
 722 AGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGGCCCACTGGGGATAT 781
 162968 AGCCCACTATGAGATATGAGGCCCACTCCAGATACGGAACCTATATGAGATC 163027
 782 TGAAGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGGCCCACTGGAGG 841
 163028 CGGTTCCTCCCACTCCAGATATGAGGCCCGGCTATGAGGATCTCTCTCAAGG 163087
 842 ATATGAGGCCCACTCCAGATATGAGGCCCACTCCAGATATGAGGCCCACTCTGC 901
 163088 GCGTAGTTCATCTCTCTGATCAAGGGCCAGCTGTATGACACAGAAAGCTCACTGC 163147
 902 TGAATAATAGAGTCTCTCAAGATCTATGACAGTCAAGCTCAAGCTCAAGCTTC 952
 163148 TGGGTCTGAAGAGGCGATCCATGTCTGTGGCAGTCCGAATCTGAATTCAGGCAATC 163207
 953 TCTTCCCACTCACTATCTTCTTGAAGTCAATTACCACTTCTCAAGATTAACCTTGA 1012
 163208 TTTTCCCTCTACCTGCTCTCAAGTCACTCTCCGCTCTGAAGATTAACCTTGA 163267
 1013 GACTCAGCAAGC-AAAGGGCACTTAAGATGACAGTCAAGTCAAGTCAAGTCAAGT 1071
 163268 GTTTCACCAAGCAAAAGCAGCTTAATGTGAAAGTCAAGTCAAGTCAAGTCAAGT 163327

RESULT 11
 AC107527/c 253149 bp DNA 1linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-92M24, WORKING DRAFT SEQUENCE, 3
DEFINITION Unordered pieces.
ACCESSION AC107527
VERSION AC107527.5 GI:30580771
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE
ORGANISM Rattus norvegicus (Norway rat)
REFERENCE 1 (bases 1 to 253149)
AUTHORS Munzy, D., Martle, M., Metzger, M., Lee, A., Adams, C., Alder, J., Allen, C., Allen, H., Alebrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayoyagi, A., Ayodeji, M., Bacc, B., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnes, M., Benahmed, F., Bivaleto, K., Blair, J., Blankensbury, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, R., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Chen, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Diya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Dublin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Faller, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C., Gabisi, A., Genta, R., Garcia, A., Garner, T., Garza, M., Georgegeorgis, B., Geer, R., Gill, R., Grady, M., Guerra, B., Guevara, M., Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., Hodge, A., Hoggins, M., Hollins, B., Howell, S., Hulky, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kovals, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, Y., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewala, L., Louisedge, H., Lozada, R., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmud, M., Malloy, K., Mangum, A., Mangum, B., Mapue, P., Martin, K., Martin, R., Martinez, B., Marwinsky, S., McLeod, M., McNeill, T., Meenen, B., Miosavljevic, A., Miner, G., Minj, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L., Narkervic, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwem, O., Ohnomu, G., Olanmuogson, A., Pal, S., Parke, K., Pasternak, S., Paul, H., Perez, A., Perez, J., Pflanz, C., Plopper, F., Polndexter, A., Popovic, D., Primus, B., Pu, L., Puzo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Slason, I., Sitter, C., D., Smaj, D., Sneed, A., Sodergren, B., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejo, Z., Uemami, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleciyk, R., Wodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, D., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, S., Zhou, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D., Holt, R., A., Smith, H., O., Weinrock, G., and Gibbs, R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 253149)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 253149)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:23664681.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

In the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GKFL

Center clone name: CH230-92M24

Assembly program: Atlas 3.0:

Consensus quality: 242817 bases at least Q40

Consensus quality: 245517 bases at least Q30

Consensus quality: 247516 bases at least Q20

Estimated insert size: 257092; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

----- Summary Statistics

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This sequence may represent more than one clone.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 250781: contig of 250781 bp in length
 * 250782 250881: gap of unknown length
 * 250882 251966: contig of 1085 bp in length
 * 251967 252066: gap of unknown length
 * 252067 253149: contig of 1083 bp in length.

FEATURES

source

1.253149

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-92M24"

2794..3331

/note="clone_boundary"

clone_end:T7

site:EcORI

end_sequence:BH296484"

2825..3625

/note="clone_boundary"

clone_end:T7

site:EcORI

end_sequence:BH296484"

complement(246782..247500)

/note="clone_boundary"

clone_end:Sp6

site:EcORI

end_sequence:BH296486"

61494 a 61494 c 61630 g 62235 t 3750 others

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 62.9%; Pred. No. 7.1e-32; Length 253149;

Matches 304; Conservative 0; Mismatches 169; Indels 10; Gaps 2;

Qy 602 ACCAGGGAATATGAACTCCACCAAGATATGAGCCCAACCAAGGAGATATGAGC 661
 Db 117968 AGCTGCGAGATATGATCCCTCTCTCTATAGTATGATACCCCTATGAGGCTATGAGT 117909

Qy 662 CCCACTATATGAGATATGAGCCCGCTGTGAGATATGAGTCCCACTTGGGGATATG 721
 Db 117908 CCCACTCTGTGATATGAGCCCGCACTGTGATATGAGATCCCACTCTGTGATATG 117849
 Qy 722 AGTCCCACTGGGGATATGAGTCCCACTTGGGGATATGAGCCCACTTGGGGATAT 781
 Db 117848 ACCCCCACTATGAGATATGAGGCGCCCACTCTCAATATGATATGAGATAT 117789
 Qy 782 TGGATGCCAAGCTGGGGATATGAGCCCGCACTTGGGGATATGAGCCCACTTGG 841
 Db 117788 CGATTCCTCCACTCCCGATATGAGCCCGCACTTGGGGATATGAGCCCGCTCA 117729
 Qy 842 ATATGAGCCCAAGCAAGCTGTAAGATATGAGCCCACTTGGGGATATGAGTCTCA 901
 Db 117728 GAGTATGATATATATCTTGTGATATGAGATCCCTGTGTGAGCAAGAGCACT 117669
 Qy 902 TGGAAATATACGCTGCTCTTACAGATATATGAGCACTGACCA-----GAGA 952
 Db 117668 TGGGTCTGAAGAGGCAATCCCACTGTGTGAGCAAGAGCAAGATTTCAAGCA 117609
 Qy 953 TCTTCCCACTTACTATCTTTTGAAGTCAATTTACCACTTTTCAAGATTAACCTTGA 1012
 Db 117608 TTTCCCTCTACTCTGTGTGAGCAAGTCCCACTTCTCCCGTGAATGAATTAAC 117549
 Qy 1013 GACTCAGCAAGC-AAAGGCACTTAAATGAGTCAAGTCAAGTGAAGAGCAAGCAAG 1071
 Db 117548 GTTTCACCAAGCAAAACGCACTTATGATATGAGATATGAGCAAGCAAGCA 117489
 Qy 1072 GCC 1074
 Db 117488 GTC 117486

RESULT 12
 AC132969 270171 bp DNA linear HTG 20-NOV-2002
 LOCUS Rattus norvegicus clone CH230-327L20, WORKING DRAFT SEQUENCE.
 AC132969
 AC132969.3 GI:25319203
 VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 270171)
 Muzny,D,Marie,,Metzker,M,Lea,,Abramson,S,,Adams,C,,Alder,J,,
 Allen,C,,Allen,H,,Albrooks,S,,Amin,A,,Anguiano,D,,
 Anylebech,V,,Aoyagi,A,,Ayodeji,M,,Baca,R,,Baden,H,,
 Baldwin,D,,Bandaranaike,D,,Barber,M,,Barnstead,M,,Benahmed,F,,
 Biswalio,K,,Blair,J,,Blankenburg,K,,Blyth,P,,Brown,M,,
 Bryant,N,,Bunay,C,,Burch,P,,Burrell,K,,Calderton,B,,
 Cardenas,V,,Carter,K,,Cavazos,I,,Cesari,J,,Center,A,,
 Chacko,J,,Chavez,D,,Chen,G,,Chen,R,,Chen,Y,,Chen,Z,,Chu,J,,
 Cleveland,C,,Cockrell,R,,Cox,C,,Coyler,M,,Crete,A,,D'Souza,L,,
 Davila,M,L,,Davis,C,,Davy-Carroll,L,,De Anda,C,,Dederich,D,,
 Delgado,O,,Denson,S,,Deramo,C,,Ding,Y,,Dim,H,,Divya,K,,
 Draper,H,,Dugan-Rocha,S,,Dunn,A,,Durbin,K,,Duval,B,,Baves,K,,
 Egan,A,,Escotto,M,,Eugene,C,,Evans,C,A,,Falls,T,,Fan,G,,
 Fernandez,S,,Finley,M,,Flagg,N,,Forbes,L,,Foster,M,,Foster,P,,
 Frazer,C,M,,Gabriel,A,,Ganta,R,,Garcia,A,,Garner,T,,Garza,W,,
 Gebregeorgis,B,,Geet,K,,Gill,R,,Garday,M,,Guerre,M,,Guevara,W,,
 Gunaratne,P,,Haaland,W,,Hamill,C,,Hamilton,C,,Hamilton,K,,
 Harvey,Y,,Havlak,P,,Hawes,A,,Henderson,N,,Hernandez,J,,
 Hernandez,R,,Hines,S,,Hladun,S,L,,Hodgson,A,,Hogues,M,,
 Hollins,B,,Howells,S,,Huliyk,S,,Hume,J,,Idlebird,D,,Jackson,A,,
 Jackson,L,,Jacob,L,,Jiang,H,,Johnson,B,,Johnson,R,,Jolivet,A,,
 Karpathy,S,,Kelly,S,,Kelly,S,,Khan,Z,,King,L,,Kovar,C,,
 Kowals,C,,Kraft,C,L,,Lebow,H,,Levan,J,,Lewis,L,,Li,Z,,Liu,D,,
 Liu,Y,,Liu,W,,Liu,Y,,London,P,,Longacre,S,,Lopez,J,,
 Lorensheewa,L,,Louieged,H,,Lozado,R,J,,Lu,X,,Ma,J,,
 Maheshwari,M,,Mahindartne,M,,Mahmoud,M,,Malloy,K,,Mangum,A,,
 Mangum,B,,Mapua,P,,Martin,K,,Martin,R,,Martinez,E,,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Pastermak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C., Plopper, F., Polidexter, A., Popovic, D., Pflums, B., Pu, L.-L., Puzos, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R., Rively, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.O., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, B., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Suton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umant, K., Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Williams, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
Journal
Unpublished

REFERENCE
2 (bases 1 to 270171)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (05-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE
3 (bases 1 to 270171)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT
On Nov 20, 2002 this sequence version replaced gi:23908490.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
Project name: K80D
Center name: CH230-327L20

----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 251525 bases at least Q40
Consensus quality: 254049 bases at least Q30
Consensus quality: 255599 bases at least Q20
Estimated insert size: 259926; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have

* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 270171: contig of 270171 bp in length.
Location/Qualifiers
1..270171
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-327L20"
1..2686
/note="igs end extension
clone end:5p6"
misc_feature
4191..5073
/note="clone boundary
clone end:5p6
site:
end_sequence:82160678"
complement(1267941..268827)
/note="clone boundary
clone end:t7
site:
end_sequence:82160655"
BASF COUNT 67733 a 63574 c 62822 g 63249 t 12793 others
ORIGIN

Query Match 12.2% Score 172.6; DB 2; Length 270171;
Best Local Similarity 62.9%; Pred. No. 7.1e-32;
Matches 304; Conservative 0; Mismatches 169; Indels 10; Gaps 2;

QY	602	ACGAGGGAATATGGAAGTCCACGAAAGATATGAGGCCACAGGAGGATATGAGC	661
DB	15216	AGTCGCGAATATGATATCCCTCTCTATATGATATGATATGAGTATGAGT	15157
QY	662	CCACCTATGGAATATGAGGCCGCGCTGATATGAGATCCACCTGAGGATATG	721
DB	15156	CCACCTCTGAGATATGAGCCCACTGATATGAGATCCCACTCTGATATGAG	15097
QY	722	AGTCCCACTGAGGATATGAGATCCCACTGAGGATATGAGATCCCACTGAGGAT	781
DB	15096	AGCCCACTATGAGATATGAGCCCACTGATATGAGATCCCACTCTGATATGAG	15037
QY	782	TGAGATCCCACTGAGGATATGATGCCCCCACTGAGGATATGAGGCCCACTGAG	841
DB	15036	CGTTCCTCCCACTCCCGATATGAGCCCGCTATATGAGTATGAGATCCCTGAG	14977
QY	842	ATATGAGGCCCACTGAGTATGAGATCCCGCTATATGAGTATGAGTATGAGT	901
DB	14976	GATGATGTCATATCTTATGATCCCACTCTGATATGAGATCCCACTGATATGAG	14917
QY	902	TGGAATATGAGTCTCTCAAGATATGAGATGAGTATGAGTATGAGTATGAGT	952
DB	14916	TGGATGATATGAGATATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT	14857
QY	953	TCTTCCCACTATCTTATCTTATGATGATATGAGTATGAGTATGAGTATGAGT	1012
DB	14856	TTTCCCTCTATCTCTGATGATATGAGTATGAGTATGAGTATGAGTATGAGT	14797
QY	1013	GATCAGCAAGC-AAAAGGCAAGTATGAGTATGAGTATGAGTATGAGTATGAGT	1071
DB	14796	GTTTCAGCAAGCAAGCAAGCAGTATGAGTATGAGTATGAGTATGAGTATGAGT	14737
QY	1072	GCC 1074	
DB	14736	GTC 14734	

RESULT 13
AF499026 793 bp mRNA linear ROD 23-APR-2002
LOCUS AF499026
DEFINITION Rattus norvegicus Wt-domain binding protein 2 (Wbp2) mRNA, complete
CDB.

```

ACCESSION      AF499026
VERSION        AF499026.1  GI:20263980
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 793)
Zanmini, M. and Nitsch, R.
Isolation of the cDNA encoding for rat Wbp2
Unpublished
2 (bases 1 to 793)
Zanmini, M. and Nitsch, R.
Direct Submission
Submitted (04-Apr-2002) DBPOM, University of Naples Federico II,
via Pansini 5, Naples, NA 80131, Italy
Location/Qualifiers
1..793
/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
1..793
/gene="Wbp2"
1..789
/contig=Wbp2

```

```

/codon_start=1
/product="Ww-domain binding protein 2"
/protein_id="AAM18132.1"
/db_xref="gi:20263981"
/tranlation="MALNKNHSGGAVYNNNTSILMSYDHELFPNDKQVPEAPKAF
TKGKVVYLLPYRVFLSKGDKAMRFPMPFPYLMKDCVQAPVGNPFIKGTVKALAGQ
MGESASVYKTLFTACGATLRFQGMQLQVLSLRGVNCAVGPVPMESGAYVFPFPPV
NGMTSGCPQYIPFPFPPPPYRFPQPMDDAMGCVQGGPPPPYRPPMHPMPPVSGSAPFT
AAEAALAAALASAVYINPFGNPNHNVITSDQPPYIPFEDKKTQ"
BASE COUNT      189 a      196 g      149 c
ORIGIN

```

Query Match	Similarity	9.1%;	Score 128.4;	DB 10;	Length 793;
Best Local	Similarity	60.3%;	Pred. No. 7.5e-21;		
Matches	251;	Conservative	0;	Mismatches	156;
				Indels	9;
				Gaps	2

QY	36	ATGCGAGTGAACCAAGACCAACCCAGAGCCGTCGTGGGGCCCTCATCCCTCTGGCCGA	95
DB	1	ATGGCGCTCAACAAAGAAATCACTCAGAGGGCGGGGAGTATGTCACAACT--GAG	57
QY	96	AGTGTCTGAACAGGTGAGGATGTGACCTGTGCTCTTACAGAAACCAATGGAATCC	155
DB	58	AGCATCTTAATGTCCTATGATCATGTGGAACCTTACATTAACAGACATGAAGATATCCA	117
QY	156	TATCTCTTTAATGACCAAGAAAGAAAGCGTTGTTCTCACTTCATACCGAGTGGTCTTC	215
DB	118	GAGGCGCTTCAAGGGGACCAAGAAAGCAACGTCTACCTAATCTCAATACCGGGTCATCTTT	177
QY	216	GTGACTTACACTTATGTCAATGACCCCATGCTTTCTTTATATGATGCCGTTGGCTGTAG	275
DB	178	CTGTC-----CAAGGGCAAGGATGTCATGCGGTCCTCATATATGCCCTTCTCAACCTGAG	231
QY	276	AGTGACTGCACATTGGAACAAACCAATTTTGGCCCCCACTACATTAAAGAAACCAATTGAG	335
DB	232	AAGGACTGTGAAGTCAAGCAACGCCGGTGTGTTGGTCCAACTTCAATTAAGGGGACATGGAAA	291
QY	336	GCAGCTCCAGGTGGTGGCTGGGAAGACAAGCTGTTTTTAAGTTATCTTCAAGAAAGGA	395
DB	292	GCTGAAGCAGGAGGTGGTGGGAAGGCTGTGCTCTCTTAACAACTGACTTTCACAGCAGGG	351
QY	396	GGTGCATCGAATTTGGCCCACTGAATGTGAAAAGCTGCGCTGCTGCTGCTGCGCAGAG	451
DB	352	GGTGCATTTAGTTGGGACGATGCTCCACAGTGGGACATCTCAAGCTTCCAGG	407

RESULT 14
MMU40826

LOCUS	MW40826	1802 bp	mRNA	linear	ROD 04-WAR-2003
DEFINITION	Mus musculus WW-domain binding protein 2 mRNA, complete cds.				
ACCESSION	U40826				
VERSION	U40826.1	GI:1777578			
KEYWORDS					
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCES
AUTHORS
TITLE

1 (sites)
Sudol, M., Chen, H. I., Bouveret, C., Rindone, A. and Bork, P.
Characterization of a novel protein-binding module--the WW domain
PERS Lett. 369 (1), 67-71 (1995)
95369475
7641887
2 (bases 1 to 1802)
Chen, H. I. and Sudol, M.
The WW domain of Yes-associated protein binds a proline-rich ligand

JOURNAL
MEDLINE
FOUNDER
REFERENCE
AUTHORS
TITLES
JOURNAL

3-binding modules
Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7819-7823 (1995)
95972270
7644438
3 (bases 1 to 1802)
Chen, J. and Sudol, M.
Identification and Characterization of Protein Ligands to the WW
Domain by Western Ligand Blotting
(in) Marshak, D.R. (Ed.);

REFERENCE
AUTHORS
TITLE
JOURNAL

Academic Press, Inc., USA (1996)
4 (bases 1 to 1802)
Chen, H.I. and Sudol, M.
Direct Submission
Submitted (17-NOV-1995) Henry I. Chen, Laboratory of Molecular
Oncology, The Rockefeller University, 1230 York Avenue, New York,
NY 10021, USA

```

source
1..1802
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/dev_stage="16 day embryo"
40..825
/note="WB-2; putative protein ligand to the WW domain of
yes-kinase associated protein YAP"
/codon_start=1
/product="WW-domain binding protein 2"
/protein_id="AAB0893.1"
/db_xref="GI:177579"
/translacion="MALNRKHSBGQYIVNNTESILMSYDHVELTFNDKRVNPAFKQ
TKKCTVILPRVIFLSKQDAMQSPMPPFLMKDCRIKQVPGANPIKGIIVKABAGG
GMBESASVYKLTFTAGCAIFBQGRNLQVASQSRBVRNGAIGYFMSGATVFPVPP
NGMVPFGPYVPPPPPEPFYGGPMDMGAKCYQVPPPPYGGPMBPVSGSADATPPTP
AARAAARAAASAVYNPCNPNNVWPTSGQPPPPYPPPPDKKTQ"

BASE COUNT      396 a      565 c      437 g      404 t
ORIGIN

```

Query Match	9.14	Score 128.4	DB 10	Length 1802
Best Local Similarity	60.34	Pred. No. 7.4e-21		
Matches 251	Conservative	0	Mismatches 156	Indels 9
			Gaps	2
QY	36	ATGGCAGTGAACCAAGAGCCACCCGAGGCGCGTGTGTGTGGGCGCTCATCCCTCTGGCCAA	95	
DB	40	ATGGCGCTCAACCAAGAAATCACTCAGAGGGGGCGCGAGTATGTCTCAACAACT---	GAG 96	
QY	96	AGTGTCTTGAAGCAGTGTGAGATGTGACCTCTGCTTCTTACAGAAAACAATGGAATCC	155	
DB	97	AGCATCTCAATGTCTTATGATCATGTGAGGCTTACCTTCAACGACATGAAATATGTGCCA	156	
QY	156	TATCTCTTAAATGCGACAAAGAAAGAAAGCTTGTCTTCTCATCTGATCCGGGTGTCTTC	215	
DB	157	GAGCGCTTCAAAAGGACCAAGAAAGAGCCGCTTACTTACTTCCGTAACGGGTATCTTT	216	

Qy 216 GTGACTTCACTTAGTCAATGACCCAGCTTCTTTATGATGCCGTTGGCCGTGATG 275
Db 217 CTGTCTAAG-----GGAGAGGAGCCCATGAGTCTTCTATGATGCCCTTTTACTGATG 270
Qy 276 AGTACTGACATCTGAAACACCAATTTTGGCCCAACTATCTTAAAGAACATCTAG 335
Db 271 AAGGACTGTGAGATTAGACAGCCGGTGTGTGTGAACTTATTAAGGATAGTGAAG 330
Qy 336 GCAAGTCAGGATGAGTGGGAGAGACAGCTGTTTATGTTATCTTCAAGAAAGGA 395
Db 331 GCTAGAGCAGGAGGGCTGGAGAGCTCCGCTCTTCAAGCTACCTTCAAGACAGAG 390
Qy 396 GATGCGCATGCAATTTGCCCACTGATGTTAAAGCTGCTGCTGACGACAGG 451
Db 391 GGGCGCATGAGTTTGGGAGAGATGCTCCAGAGTGCATCTCAAGCTCCAGAGG 446

RESULT 15
BX470149/c
LOCUS
DEFINITION BX470149 222469 bp DNA linear HTG 07-MAY-2003
Dario rerio clone DKEX-242K7, *** SEQUENCING IN PROGRESS ***, 50
unordered pieces.
ACCESSION BX470149
VERSION BX470149.3 GI:30424228
KEYWORDS HTG; HTGS PHASE1.
SOURCE Dario rerio (zebrafish)
ORGANISM Dario rerio
Bukarjota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 222469)
REFERENCE
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zf1sh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On May 7, 2003 this sequence version replaced gi:30387077.
COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zf1sh-help@sanger.ac.uk
----- Project Information
Center project name: ZK242K7
----- Summary Statistics
Assembly program: XGAP; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 197958 bases at least Q40
Consensus quality: 206715 bases at least Q30
Consensus quality: 212351 bases at least Q20
Insert size: 217569; sum-of-contigs
Insert size: 165837; 6.2% error; agarose-fp
Quality coverage: 2.48x in Q20 bases; sum-of-contigs Quality
coverage: 3.95x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will
* be preserved.
*
* 1
* 6656: contig of 6656 bp in length
* 6657 gap of 100 bp
* 6757 10255: contig of 3499 bp in length
* 10256 10355: gap of 100 bp
* 10356 13223: contig of 2868 bp in length
* 13224 13323: gap of 100 bp
* 13324 19935: contig of 6612 bp in length
* 19936 20035: gap of 100 bp
* 20036 22597: contig of 2562 bp in length
* 22598 22697: gap of 100 bp

* 22698 28569: contig of 5872 bp in length
* 28570 28669: gap of 100 bp
* 28670 37965: contig of 9296 bp in length
* 37966 38065: gap of 100 bp
* 38066 41140: contig of 3075 bp in length
* 41141 41240: gap of 100 bp
* 41241 45108: contig of 3868 bp in length
* 45109 45208: gap of 100 bp
* 45209 47730: contig of 2522 bp in length
* 47731 47830: gap of 100 bp
* 47831 51884: contig of 4054 bp in length
* 51885 51984: gap of 100 bp
* 51985 55857: contig of 3873 bp in length
* 55858 55957: gap of 100 bp
* 55959 60498: contig of 4541 bp in length
* 60499 60598: gap of 100 bp
* 60599 63641: contig of 3043 bp in length
* 63642 63741: gap of 100 bp
* 63742 72674: contig of 8933 bp in length
* 72675 72774: gap of 100 bp
* 72775 75602: contig of 2828 bp in length
* 75603 75702: gap of 100 bp
* 75703 78062: contig of 2360 bp in length
* 78063 78162: gap of 100 bp
* 78163 84485: contig of 6323 bp in length
* 84486 84585: gap of 100 bp
* 84586 87553: contig of 2968 bp in length
* 87554 87654: gap of 100 bp
* 87654 92487: contig of 4834 bp in length
* 92488 92587: gap of 100 bp
* 92588 96104: contig of 3517 bp in length
* 96105 96204: gap of 100 bp
* 96205 101479: contig of 5275 bp in length
* 101480 101579: gap of 100 bp
* 101580 104979: contig of 3400 bp in length
* 104980 105079: gap of 100 bp
* 105080 108003: contig of 2924 bp in length
* 108004 108103: gap of 100 bp
* 108104 113151: contig of 5048 bp in length
* 113152 113251: gap of 100 bp
* 113252 116585: contig of 3334 bp in length
* 116586 116685: gap of 100 bp
* 116686 119345: contig of 2661 bp in length
* 119347 119445: gap of 100 bp
* 119447 123125: contig of 3679 bp in length
* 123126 123225: gap of 100 bp
* 123226 131266: contig of 8041 bp in length
* 131267 131366: gap of 100 bp
* 131367 136193: contig of 4827 bp in length
* 136194 136293: gap of 100 bp
* 136294 139482: contig of 3189 bp in length
* 139483 139582: gap of 100 bp
* 139583 141875: contig of 2293 bp in length
* 141876 141975: gap of 100 bp
* 141976 144123: contig of 2147 bp in length
* 144123 144222: gap of 100 bp
* 144223 146968: contig of 2746 bp in length
* 146969 147068: gap of 100 bp
* 147069 149399: contig of 2331 bp in length
* 149400 149499: gap of 100 bp
* 149500 153963: contig of 4464 bp in length
* 153964 154063: gap of 100 bp
* 154064 156579: contig of 2516 bp in length
* 156580 156679: gap of 100 bp
* 156680 158928: contig of 2249 bp in length
* 158929 159028: gap of 100 bp
* 159029 164542: contig of 5514 bp in length
* 164543 164642: gap of 100 bp
* 164643 172054: contig of 7412 bp in length
* 172055 175210: contig of 3056 bp in length
* 175211 175310: gap of 100 bp
* 175311 179379: contig of 4069 bp in length

	*	179380	179479:	gap of 100 bp	
	*	179480	182698:	contig of 3219 bp	in length
	*	182699	182798:	gap of 100 bp	
	*	182799	186779:	contig of 3581 bp	in length
	*	186780	186879:	gap of 100 bp	
	*	186880	195626:	contig of 8747 bp	in length
	*	195627	195726:	gap of 100 bp	
	*	195727	199396:	contig of 3670 bp	in length
	*	199397	199496:	gap of 100 bp	
	*	199497	206412:	contig of 6316 bp	in length
	*	206413	206512:	gap of 100 bp	
	*	206513	209619:	contig of 3107 bp	in length
	*	209620	209719:	gap of 100 bp	
	*	209720	219083:	contig of 9364 bp	in length
	*	219084	219183:	gap of 100 bp	
	*	219184	222469:	contig of 3286 bp	in length
FEATURES					
Source			Location/Qualifiers		
			1..222469		
			/organism="Danto rerio"		
			/mol_type="genomic DNA"		
			/db_xref="taxon:7955"		
			/clone="DXEX-242KX"		
			/clone_1lb="Dantokex"		
misc_feature			1..6656		
			/note="assembly_fragment:01359		
misc_feature			fragment_chain:1"		
			6757..10255		
			/note="assembly_fragment:01385		
misc_feature			fragment_chain:1"		
			10356..13223		
			/note="assembly_fragment:00592		
misc_feature			fragment_chain:1"		
			13324..19935		
			/note="assembly_fragment:00237		
misc_feature			fragment_chain:1"		
			20036..22597		
			/note="assembly_fragment:00949		
misc_feature			fragment_chain:1"		
			22698..28569		
			/note="assembly_fragment:00615		
misc_feature			fragment_chain:1"		
			28670..37965		
			/note="assembly_fragment:01306		
misc_feature			fragment_chain:2"		
			38066..41140		
			/note="assembly_fragment:00487		
misc_feature			fragment_chain:2"		
			41241..45108		
			/note="assembly_fragment:00232		
misc_feature			fragment_chain:2"		
			45209..47720		
			/note="assembly_fragment:00935		
misc_feature			fragment_chain:2"		
			47831..51864		
			/note="assembly_fragment:01415		
misc_feature			fragment_chain:2"		
			51985..55857		
			/note="assembly_fragment:00878		
misc_feature			fragment_chain:3"		
			55958..60498		
			/note="assembly_fragment:01208		
misc_feature			fragment_chain:3"		
			60599..63641		
			/note="assembly_fragment:01266		
misc_feature			fragment_chain:3"		
			63742..72674		
			/note="assembly_fragment:01329		
misc_feature			fragment_chain:3"		
			72775..75602		
			/note="assembly_fragment:01183		
misc_feature			fragment_chain:4"		
			75703..78062		
			/note="assembly_fragment:01179		

```

Query Match      8.7%; Score 122.8; DB 2; Length 222469;
Best Local Similarity 88.7%; Prod. No. 1,6e-19;
Matches 133; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      206  GGTGGTCTTCGGGACTTCACACTTAATGTCATGACCCCAAGCTTCTTTATGATGCCGTT 265
Db       190751  GGGGAATCTTCGGGACTTCACACTGATGAGACCCCAAGTTCCTTCATGAGGCATT 190692

QY      266  TGGCTGATGAGTGACTGCACCATTAACAACCAATTTTSCCCCAACATCATTAAGA 325
Db       190691  TGATCTGATGAGTAACTGCACCATTAACAACCGCTCTTSCCCCAACATCATTAAG 190633

QY      326  AACCATTCAGGCACTCCAGGTGGTGGCTG 355
Db       190631  AACCGTTAGGCACTCCAGATGGTGAATG 190602

Search completed: December 16, 2003, 15:10:06
Job time : 5266.22 secs

```

Search completed: December 16, 2003, 15:10:06
Job time : 5266.22 secs


```

FT      given in AAS20604"
XX      WO200190185-A2.
XX      29-NOV-2001.
XX      25-MAY-2001; 2001WO-CA00738.
XX      25-MAY-2000; 2000CA-2307128.
XX      25-MAY-2000; 2000US-206979P.
XX      (TOOH ) UNIV QUEBENS KINGSTON.
XX      (UYOR-) UNIV OREGON HEALTH SCI.
XX      Oko R, Sutovsky P;
XX      WPI, 2002-097644/13.
XX      P-PSDB; AAU74604.
XX      Isolated perinuclear theca 32 polypeptide that interacts with activated
XX      tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing
XX      diminished fertility and abnormal spermiogenesis and for providing
XX      contraception -
XX      Claim 10; Fig 2A-B; 103pp; English.
XX      The invention describes an isolated perinuclear theca 32 (PT32)
XX      polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is
XX      useful for: enhancing fertility in a mammal; treating globozoosperm, by
XX      expressing (I) in spermatozoa; inhibiting fertilisation, by introducing
XX      (I) or its antigenic fragment into a mammal to elicit an immune
XX      response; enhancing the ability of round spermatids to activate oocytes;
XX      treating or diagnosing diminished fertility and abnormal spermiogenesis;
XX      in providing contraception; identifying contraceptive and
XX      fertility-enhancing agents. The polynucleotide is useful for producing
XX      (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
XX      for chromosome identification. An antibody against (I) is useful in
XX      immunological assays. In immunoneutralization methods, to identify cells
XX      expressing (I), and to purify (I) by affinity chromatography. A
XX      transgenic animal is useful as an animal model for studying human
XX      fertility and reproductive biology, and for screening compounds to
XX      identify modulators of oocyte activation. The use of (I) prevents the
XX      entry of components which are detrimental to embryonic development into
XX      the oocyte during oocyte activation with crude sperm extract and avoids
XX      the propagation of viruses such as HIV (human immunodeficiency virus) and
XX      SIV (Sialian immunodeficiency virus) carried in the sperm. This sequence
XX      encodes the bovine perinuclear theca 32 (PT32), described in the method
XX      of the invention.
XX      Sequence 1413 BP; 377 A; 363 C; 369 G; 304 T; 0 other;
SQ
Query Match      100.0%; Score 1413; DB 24; Length 1413;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1413; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY      1 CGGCACGAGGGCGCGCAGAGAGGGGCGCTGGGCGAGATGCGAGTGAACAGAGCCACCG 60
DB      1 CGGCACGAGAGGGCGCGCAGAGAGGGGCGCTGGGCGAGATGCGAGTGAACAGAGCCACCG 60
QY      61 AGAGCCGTCGTGGGGCCCTCATCCCTCTGGCGAAAGTGTCTTGAAGCAGTGTGAGATG 120
DB      61 AGAGCCGTCGTGGGGCCCTCATCCCTCTGGCGAAAGTGTCTTGAAGCAGTGTGAGATG 120
QY      121 TGGACCTCTGCTCCGACAGAAACAGAGTGAATCCATCTCTTTATGGCAGAAAGAAAG 180
DB      121 TGGACCTCTGCTCCGACAGAAACAGAGTGAATCCATCTCTTTATGGCAGAAAGAAAG 180
QY      181 GAAAGCTGTCTTCACTTCATACCGAGGTGTCTTGTGATCTTCACTTAACTGCAATGACC 240
DB      181 GAAAGCTGTCTTCACTTCATACCGAGGTGTCTTGTGATCTTCACTTAACTGCAATGACC 240
QY      241 CCAATGCTTTCTTTATGATGCCGTTTGGCTGATGATGATGACATGCACTTGAACCAACCA 300

```

Db	241	CGAAGCTTTCTTTATATGATGCGGTTTGCCCTGATGATGATCTGCAACATTTGAAACAACCA	300
Qy	301	TTTTTGGCCCCCACTAATTAATTAAGAAACCAATTGACGGACGCTCCAGGTGGTGGCTGGAG	360
Db	301	TTTTTGGCCCCCACTAATTAATTAAGAAACCAATTGACGGACGCTCCAGGTGGTGGCTGGAG	360
Qy	361	GACAAAGCTGTTTTTAAGTTATCCCTTCAGAAAGGAAGGTGCCATGGAATTGGCCCAACTGA	420
Db	361	GACAAAGCTGTTTTTAAGTTATCCCTTCAGAAAGGAAGGTGCCATGGAATTGGCCCAACTGA	420
Qy	421	TGGTAAAGCTGCTCTGCTGCTGCGCAGAGGAATTCACATTGGAAGGTAAATTACTGGT	480
Db	421	TGGTAAAGCTGCTCTGCTGCTGCGCAGAGGAATTCACATTGGAAGGTAAATTACTGGT	480
Qy	481	TGCAACTCTTCAGAGACTGATCAATAATTACTGTCCAGAGGGCTGCAAGTGTGCTCTCAACA	540
Db	481	TGCAACTCTTCAGAGACTGATCAATAATTACTGTCCAGAGGGCTGCAAGTGTGCTCTCAACA	540
Qy	541	CACCTGTGCTCCAGCATATCAATTGTGATCTATGAGCCGCCACCAAGCATATACAGTCC	600
Db	541	CACCTGTGCTCCAGCATATCAATTGTGATCTATGAGCCGCCACCAAGCATATACAGTCC	600
Qy	601	AACCAAGGGGAATATGAACTTCCACAGAGAATATGAGCCCAACAGAGGGGATATGAG	660
Db	601	AACCAAGGGGAATATGAACTTCCACAGAGAATATGAGCCCAACAGAGGGGATATGAG	660
Qy	661	CCCCACTATAGGGAATATGAGACCCCGCTGTGGGATATGAGTCCCACTGGGGGATATG	720
Db	661	CCCCACTATAGGGAATATGAGACCCCGCTGTGGGATATGAGTCCCACTGGGGGATATG	720
Qy	721	GAGTCCCACTGGGGGATATGAGATCCCACTGGGGGATATGAGGCCCACTGGGGGAT	780
Db	721	GAGTCCCACTGGGGGATATGAGATCCCACTGGGGGATATGAGGCCCACTGGGGGAT	780
Qy	781	ATGGAAGTCCCACTGGGGGATATATGTATGCCCACTGGGGGATATGAGAGCCCACTGAG	840
Db	781	ATGGAAGTCCCACTGGGGGATATATGTATGCCCACTGGGGGATATGAGAGCCCACTGAG	840
Qy	841	GATATGAGGCCCACTGAGTAATGAAAGCCTTACCCTTCATATGAACTCATCTG	900
Db	841	GATATGAGGCCCACTGAGTAATGAAAGCCTTACCCTTCATATGAACTCATCTG	900
Qy	901	CTGGAATATAGAGCTGTCTCTCAAGATCTATATGACGTCAGCAGAGAACTTCTTTCCA	960
Db	901	CTGGAATATAGAGCTGTCTCTCAAGATCTATATGACGTCAGCAGAGAACTTCTTTCCA	960
Qy	961	CTACCTCATCTCTTAAAGTCAATTTACACCTTCTCAGAGTTAACTTGAAGTCAAC	1020
Db	961	CTACCTCATCTCTTAAAGTCAATTTACACCTTCTCAGAGTTAACTTGAAGTCAAC	1020
Qy	1021	AAGCAAGGAGCACTTAAACTGAAGTCAAGTAAAGAGAGAACCAAGGTGCCAGTGG	1080
Db	1021	AAGCAAGGAGCACTTAAACTGAAGTCAAGTAAAGAGAGAACCAAGGTGCCAGTGG	1080
Qy	1081	TAGAGAGGTGTTGTTGTGTGACGCAATGTGTCTGATCTTCTCCACAACCTGTGAGGTCTGT	1140
Db	1081	TAGAGAGGTGTTGTTGTGTGACGCAATGTGTCTGATCTTCTCCACAACCTGTGAGGTCTGT	1140
Qy	1141	GCCCTCAAAACAAGATGAAGGTGAAGAACGACTCCGTGTTCTCAAGAAAGAAAGATGCTTGA	1200
Db	1141	GCCCTCAAAACAAGATGAAGGTGAAGAACGACTCCGTGTTCTCAAGAAAGAAAGATGCTTGA	1200
Qy	1201	AAACAGACTGCAAGCCCAACTAGAGAGAGAGATGTGAAGTGTGCAATTAACAGCTTGG	1260
Db	1201	AAACAGACTGCAAGCCCAACTAGAGAGAGAGATGTGAAGTGTGCAATTAACAGCTTGG	1260
Qy	1261	GGATGAGACTGACTTCTCTTTAAGAAAACAGGCTTCTCTGCTCTGACTTGACGAGAA	1320
Db	1261	GGATGAGACTGACTTCTCTTTAAGAAAACAGGCTTCTCTGCTCTGACTTGACGAGAA	1320
Qy	1321	AAGAGAAATGCGTGGAAACAAAGAGCTAAGGATCAACCGCTTAAAGCGCCCTGATTAAG	1380
Db	1321	AAGAGAAATGCGTGGAAACAAAGAGCTAAGGATCAACCGCTTAAAGCGCCCTGATTAAG	1380

Oy		1381 CCTGCTGGTGTGCATAAAAA	1413
Dd		1381 CCGCTGCTGTTTCATTAAAAA	1413
		RESULT 2	
		AAS20602	
		ID AAS20602 standard; cDNA, 1001 BP.	
Xx			
Xc		AAS20602;	
Xx			
D7		09-APR-2002 (first entry)	
Xx			
D8		DNA encoding human testicular WW domain binding protein (htwbp).	
Xx			
Kw		Testicular WW domain binding protein; WBPH; perinuclear theca 32;	
Kw		PTJ2; contraceptive; fertility; oocyte activation; vaccine;	
Kw		globozoospermy; spermogenesis; spermatozoa; tyrosine kinase; c-Yes;	
Xx		immunoreceptive; human; gene; ss.	
Xs		Homo sapiens.	
Xk			
Fh		Key Location/Qualifiers	
Ft		primer_bind 1..18	
Ft		/tag= A	
Ft		/note= "Binds forward primer AAS20605"	
Ft		CDS 1..708	
Ft		/tag= b	
Ft		/product= "htwBP"	
Ft		/note= "Human testicular WW domain binding protein"	
Ft		primer_bind complement (984..1001)	
Ft		/tag= c	
Ft		/note= "Binds reverse primer AAS20606"	
Pn		WO200190185-A2.	
Pd		29-NOV-2001.	
Pp		25-MAY-2001; 2001WO-CA00738.	
Xx			
Xx		25-MAY-2000; 2000CA-2307128.	
Pr		25-MAY-2000; 2000US-206979P.	
Xx			
Pa		(TOOH) UNIV QUEBENS KINGSTON.	
Pa		(UYOR-) UNIV OREGON HEALTH SCI.	
Xx			
Pi		Oko R, Sutovsky P;	
Pi			
Dr		WPI; 2002-097644/13.	
Dr		P-PsDB; AAU74610.	
Xx			
Pt		Isolated perinuclear theca 32 polypeptide that interacts with activated	
Pt		tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing	
Pt		diminished fertility and abnormal spermogenesis and for providing	
Pt		contraception -	
Xx			
P8		Claim 62; Fig 4B; 103pp; English.	
Xx			
Xx		The invention describes an isolated perinuclear theca 32 (PTJ2)	
Cc		polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is	
Cc		useful for: enhancing fertility in a mammal; treating globozoospermy; by	
Cc		expressing (I) in spermatozoa; inhibiting fertilisation, by introducing	
Cc		(I) or its antigenic fragment into a mammal to elicit an immune	
Cc		response; enhancing the ability of round spermatids to activate oocytes	
Cc		treating or diagnosing diminished fertility and abnormal spermogenesis	
Cc		in providing contraception; identifying contraceptive and	
Cc		fertility-enhancing agents. The polynucleotide is useful for producing	
Cc		(I) by recombinant techniques, as vaccine, as diagnostic reagents, and	
Cc		for chromosome identification. An antibody against (I) is useful in	
Cc		immunological assays, in immunoreceptive methods, to identify cells	
Cc		expressing (I), and to purify (I) by affinity chromatography. A	
Cc		transgenic animal is useful as an animal model for studying human	

	fertility and reproductive biology, and for screening compounds to
CC	identify modulators of oocyte activation. The use of (i) prevents the
CC	entry of components which are detrimental to embryonic development into
CC	the oocyte during oocyte activation with crude sperm extract and avoids
CC	the propagation of viruses such as HIV (human immunodeficiency virus) and
CC	SIV (simian immunodeficiency virus) carried in the sperm. This sequence
CC	encodes the human testicular MW domain binding protein (htMBP), described
in	the method of the invention.
XX	
Seq	Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
Query Match	30.8%; Score 435.4; DB 24; Length 1001;
Best Local Similarity	74.7%; Pred. No. 2,4e-117;
Matches 609;	Conservative 0; Mismatches 181; Indels 25; Gaps 4;
Dy	ATGCCGTTTGGCTGTATGATGA CTGCAACATTGAACAACAAATTTTGCCCCAACATAC 317
Dy	1 ATGCATTTGATCTGATGAGCAACACTCAGCTTGGAACAACAGTAATTTGCGCAACTTC 60
Dy	318 ATTAAAGAAACCATTCAGGAGCTCCAGGTGTGTGCTGGGAAGACAAGCTGTTTTAAG 377
Dy	61 ATTAAAGGAACCTATTCAGGACGCTCCATAATGTGTGTGGGAAGACAAGCTACTTTAAA 120
Dy	378 TTATTCCTTCAGGAAGAAGGTGCGATGAAATTTTGCCCACTGATGTAAAAAGTCCCT 437
Dy	121 TTATGCTTCAGAAATGAGATGCAATTTGCCCAGTTGATGTGAAAGTCTCTCT 180
Dy	438 GCTGCTGCGAAGGAATTCACCTTGGAAGTGAATTAATTAATCTGTTCGACATTCAGAC TG 497
Dy	181 GCTGTTCGCCAGGAATTTCCACTTAAGAACTTAAATGACGTGTTCACCTGTATGGGAAT 240
Dy	498 TACATAATTAATCTGCCAGGGGCTGACAGTGTCTCTCAACAACACTTGTCCAGCATAT 557
Dy	241 TATGTATAATTCGTGGGAGGGAAT--ATGTGACCTCAACAAGTCCCTGTTCAG----- 292
Dy	558 CCAATGTGATCTATGGAACCCCAACAACAAGATATACATGCCAACAGGGGAATATGGA 617
Dy	293 ----TATGTGCTATGAGGGCCCAACTGCAAGATATGAGGCCCACTCCGATATGGA 348
Dy	618 ACTCCACAGAAGATATGAGGCCCAACAAGGGGGAATATGAGGCCCACTATGGGATAT 677
Dy	349 GCCCACCTGCAAGATATGAGGCCCAACCGTAGGAATATGAAGCCCGCTGTGGGATAC 408
Dy	678 GGAAGCCCGCTGTGGGATATGAGCTCCACTGTGGGGATATGAGTCCCACTGGGGGA 737
Dy	409 AGAGCCTCACTGTGCGATATGAGGCCCACTCTTGATATACGGAAGCCCACTGCAAGA 468
Dy	738 TATGGAATCCCACTGTGGGGATATGAGGCCCACTGTGGGGATATGAGTCCCACTGGG 797
Dy	469 TATGGAACCCCACTCTAGGATATGAGGCCCACTCTTGATATGGAACCCCACTCTC 528
Dy	798 GGATATGATGTCGCCCACTGTGGGGATATGAGGCCCACTGTGAGATATGAGGCCCAACA 857
Dy	529 GAATATGAGGCCCACTCTCGATATGAGGCCCACTGTGAGAAATATGAAGCCCGCTCT 588
Dy	858 GCTGGAATATGAGGCCCTTACCCTCTGCAATATGAAGCTCATCTGTGGAATAATACGTGC 917
Dy	589 GCGGATATACAGAGCTCACTGTGATCAAGAGCCAGGACCTCAAGGAATATCAAGCAGCC 648
Dy	918 TCTCAAGATCTATGACAGCTCAAGAGAACTTCTTCCACTTAATCTACTCTTTCTTAA 977
Dy	649 -----CAGGCTCTCTGAAAGAGAGCTCTCTCCCTCTGCTCTCTCTCTCAAG 696
Dy	978 GTGCATTTACCACTTCTCAAGATTAACCTTGAAGCTCAACAAGGAAG--GGACCTCT 1038
Dy	697 GTCAATCTTAACCTTCTTAAGATTAACCTTGAAGCTCAACAAGGAAGATACCTCT 756
Dy	1037 AAAATGAGTCAAGTAAAGGAAGAACCCAGGT 1071
Dy	757 AAAATGAGTCAAGATTAAGAGGACGACTCAGCT 791

ABK84798
ID ABK84798 standard; cDNA; 220895 BP.
AC ABK84798;
XX
XX 14-AUG-2002 (first entry)
XX
XX Human cDNA differentially expressed in granulocyte cells #1369.
DE
XX
XX
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
viral infection; parasitic infection; protozoal infection;
fungal infection; sterile inflammatory disease; psoriasis;
rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
cardiac reperfusion injury; renal reperfusion injury; ARDS;
adult respiratory distress syndrome; inflammatory bowel disease;
Crohn's disease; ulcerative colitis; periodontal disease;
granulocyte activation; chronic inflammation; allergy.
KM
KM
XX Homo sapiens.
OS
XX
XX MO200228999-A2.
PN
XX
XX 11-APR-2002.
PD
XX
XX 03-OCT-2001; 2001MO-US30821.
PF
XX
XX 03-OCT-2000; 2000US-237189P.
PR
XX
XX (GENE-) GENE LOGIC INC.
PA
XX
XX Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;
PI
XX MPI; 2002-435328/46.
DR
XX
XX Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity -
XX
XX
PS Claim 1, SEQ ID No 1369, 114pp; English.
XX
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated
CC GC, where differential expression of Gs is indicative of GCA.
CC Also included are modulating (M2) GA by contacting GC with an agent
CC that alters the expression of at least one gene in Gs; (2) screening (M3)
CC for an agent capable of modulating GCA or an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease using the
CC gene expression profile; (3) detecting (M4) an inflammation (especially
CC chronic) in a tissue, an allergic response in a subject, exposure of a
CC subject to a pathogen or sterile inflammatory disease, by detecting the
CC level of expression in a sample of the tissue of gene(s) from Gs, where
CC the level of expression of the gene is indicative of inflammation;
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC an allergic response in a subject, exposure of a subject to a pathogen
CC or sterile inflammatory disease, by contacting a tissue having
CC inflammation with an agent that modulates the expression of gene(s)
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC modulating GA; M3 is useful for screening an agent capable of modulating
CC GCA preferably in an inflammation in a tissue; M4 is useful for
CC detecting an inflammation (especially chronic) in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease, also bacterial infection, viral infection, also
CC parasitic infection, protozoal infection, fungal infection and M5 is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_jct_sequences.

8Q Sequence 220895 BP, 53760 A, 54808 C, 55831 G, 56496 T, 0 other;

Query Match	20.0%	Score 282.6	DB 24	Length 220895
Best Local Similarity	74.6%	Pred. No. 3.6e-71		
Matches 387	Conservative 0	Mismatches 119	Indels 13	Gaps 2
QY 554 ATATCCATTGATGATCTATGAGACCCCAACAGAGATATACGTCCAAACGAGGGAATA 613				
DB 169988 ATTCCTGATATGTCATATGAGGCCCACTGACAGATATGAGGCCCACTCCGGAATA 170057				
QY 614 TGAAGCTCCACAGAGAGATATGAGGCCCAACAGAGGAGATATGAGGCCCACTATGAG 673				
DB 170058 CGAGGCCCACTGACAGATATGAGGCCCAACCTGAGAAATGAGAGCCCGCTGTGG 170117				
QY 674 ATATGAGACCCCGCTGTGAGATATGAGTCCCACTGAGGAGATATGAGTCCCACTG 733				
DB 170118 ATACAGACCTCACTGTAGCATATGAGACCCCACTCTTGATATAGGAGCCCACTG 170177				
QY 734 GGGATATGAGATCCCACTGTGGGGATATGAGACCCCACTGTGGGGATATGAGTCCCACT 793				
DB 170178 AGGATATGAGGCCCACTCTGAGATATGAGACCCCACTCTTGATATGAGACCCCACT 170237				
QY 794 TGGGGATATGAGTGCACCACTGAGGGAGATATGAGAGCCCACTGACAGATATGAGACCC 853				
DB 170238 TCTCGAATATGAGAGCCCACTCTGAGATATGAGACCCCACTGACAGAAATGAGAGCC 170297				
QY 854 ACCAGCTGAAATGAGAGCCCTACCCCTGACATATGAGTCCATCTGCTGAAATAACAG 913				
DB 170298 GCTTCGAGGATACAGAGCTCTCACTGTGATACAGAGCCAGAGCTCAGAAATCTACAG 170357				
QY 914 TGCCTCTACAGATCTATGACAGCTCAGAGAGACTTCTCTTCCACTACATCTTTC 973				
DB 170358 AGCC-----CAGGCTCTGAAACAGAGGCTTCTCTTCCCTGCTCTCTTC 170405				
QY 974 TTAGGTCATTTACCACTCTCAGAGTTAAACCTTGAAGACTCACCAAGCAAG-AGCA 1032				
DB 170406 TAGGTCATTTCTTAACCTTCTTAAGATGTAACCTTGAAGACTCACCAAGCAAGAGTA 170465				
QY 1033 CCTAAAACTGAGTCAACAGTAAGAGAGAACCCAGGT 1071				
DB 170466 CCTAAAAATTGAAGTCAAGATTAAGAGAGACGACTCAGGT 170504				
RESULT 4				
AA101229				
ID AA101229 standard; cDNA, 467 BP.				
XX AA101229;				
AC				
XX 21-NOV-2001 (first entry)				
DT				
XX Human reproductive system related antigen cDNA SEQ ID NO: 1230.				
DB				
XX Human; reproductive system related antigen; reproductive system disorder;				
KW cancer; gene therapy; ss.				
XX				
OS Homo sapiens.				
XX				
PN WO200155320-A2.				
PD				
XX 02-AUG-2001.				
PF				
XX 17-JAN-2001; 2001WO-US01339.				
PR				
XX 31-JAN-2000; 2000US-0179065.				
PR				
XX 04-FEB-2000; 2000US-0180528.				
PR				
XX 24-FEB-2000; 2000US-0184664.				
PR				
XX 02-MAR-2000; 2000US-0186350.				

PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249287.
PR 17-NOV-2000; 2000US-0249289.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-025678.
PR (HUMA-) HUMAN GENOME SCI INC.
PR PA
PR XX
PR PI Rosen CA, Barash SC, Ruben SM,
PR XX WPI; 2001-46570/50.
PR DR P-PSDB; AAM95259.
PR DR
PR XX
PR PT Isolated nucleic acid molecule encoding a reproductive system antigen
PR PT is used in preventing, treating or ameliorating a medical condition -
PR XX
PR PS Claim 1; SEQ ID NO 1230; 1297bp + Sequence Listing; English.
PR CC The present invention provides the protein and coding sequences of a

CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
CC invention.

XX Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;

Query Match 18.6%; Score 262.6; DB 22; Length 467;
Best Local Similarity 76.4%; Pred. No. 1,1e-66;
Matches 365; Conservative 0; Mismatches 100; Indels 13; Gaps 3;

QY 386 CAGAAAGAGAGTGCATGGAATTTGCCAATGATGTAAAGCTGCTGCTGCTGC 445
DB 2 CAGAAATGAGAGGTCATGGAATTTGCCAATGATGTAAAGCTGCTGCTGCTGC 61
QY 446 CAGAGGAATTTCCATTTGGAATTAATTAATTAATTAATTAATTAATTAAT 505
DB 62 CCGAGGATTTCCATTTGGAATTAATTAATTAATTAATTAATTAATTAATTAAT 121
QY 506 TACTGTCGAGAGGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 565
DB 122 TACTGTCGAGAGGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 169
QY 566 GATCTATGAGACCCCAACAGATATACAGTCCACAGAGGGAATATGAACTCCACC 625
DB 170 TGTCTATGAGAGCCCACTGAGATATGAGGCCCACTCCGGAATAGAGCCCAACC 229
QY 626 AGAAGATATGAGCCCAACAGAGGATATGAGGCCCACTGATGGAATATGAGCCCC 685
DB 230 TGCAGATATGAGCCCAACAGAGGATATGAGGCCCACTGATGGAATATGAGCCCTC 289
QY 686 GCTGTGAGATATGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 745
DB 290 ACTGTGAGATATGAGTGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 349
QY 746 CCCACCTGAGAGATATGAGGCCCACTGAGGAGATATGAGGCCCACTGAGGAGATATG 804
DB 350 CCCACCTGAGAGATATGAGGCCCACTGAGGAGATATGAGGCCCACTGAGGAGATATG 409
QY 805 GTGCCCCACCTGAGGAGATATGAGGCCCACTGAGGAGATATGAGGCCCACTGAGGAG 862
DB 410 GAGCCCCACCTGAGGAGATATGAGGCCCACTGAGGAGATATGAGGCCCACTGAGGAG 467

RESULT 5
ABL96688
ID ABL96688 strand; cDNA; 467 BP.
XX
AC ABL96688;
XX
DT 21-JUN-2002 (first entry)
XX
DE Human testicular antigen encoding cDNA SEQ ID NO: 356.
XX
KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200155317-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226881.
PR 22-AUG-2000; 2000US-0226882.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232402.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0232404.
PR 14-SEP-2000; 2000US-0232405.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.

PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2000US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483232/52.
 DR
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating testicular cancer
 XX
 PS Claim 1; SEQ ID NO 356; 766pp; English.
 CC
 CC The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and

CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a cDNA of the
 CC invention.
 CC
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
 Query Match 18.6%; Score 262.6; DB 23; Length 467;
 Best local similarity 76.4%; Pred. No. 1.1e-66;
 Matches 365; Conservative 0; Mismatches 100; Indels 13; Gaps 3;
 QY 386 CAGGAAGAGGAGGCGCATGCAATTTGGCCACGATGTAAGCTGCTCTGCTGC 445
 DB 2 CAGAAATGAGAGGCGCATTAATTTGCCGATGATGTAAGCTGCTCTGCTGC 61
 QY 446 CAGGAATTCACCTTGAAGTGAATTAATTAATTAATTAATTAATTAATTAAT 505
 DB 62 CCGAGATTCCTCACTTGAAGCTTAAATGATGATGATGATGATGATGATGAT 121
 QY 506 TACTGTCCAGAGGAGGCTGCACTGTGCTCTGCAAGACACTTGTCCAGATTCAT 565
 DB 122 TACTGGGAGAGGAGAT---ATGTGCACTCAGAGATGCTTTGAG-----TTAT 169
 QY 566 GATCTATGACCCCGACACAGAGATTAAGTCCACAGAGGAGATTAAGACTCACC 625
 DB 170 TGTCTATGAGCCCGACCTGCAAGATTAAGGCCCACTCCGAGATACGAGCCCA 229
 QY 626 AGAAGATTAAGAGCCCGACAGAGGAGATTAAGAGCCCGATTAAGGAGATTAAG 685
 DB 230 TGCAGATTAAGAGCCCGACAGAGGAGATTAAGAGCCCGATTAAGGAGATTAAG 289
 QY 686 GCTGTGAGATTAAGAGTCCACCTGAGGAGATTAAGAGTCCACCTGAGGAGATTAAG 745
 DB 290 ACTGTGAGATTAAGAGCCCGACCTGAGATTAAGAGCCCGACCTGAGGAGATTAAG 349
 QY 746 CCGAAGGAGATTAAGAGTCCCGACCTGAGGAGATTAAGAGTCCCGAGGAGATTAAG 804
 DB 350 CCGAAGGAGATTAAGAGTCCCGACCTGAGGAGATTAAGAGTCCCGAGGAGATTAAG 409
 QY 805 GTGCCCGACCTGAGGAGATTAAGAGCCCGACCTGAGGAGATTAAGAGCCCGACCTG 862
 DB 410 GAGCCCGACCTGAGGAGATTAAGAGCCCGACCTGAGGAGATTAAGAGCCCGACCTG 967
 RESULT 6
 ID ABA67850 standard; DNA; 436 BP.
 XX
 AC ABA67850;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DS Human foetal liver single exon nucleic acid probe #16155.
 XX
 KF Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX
 OS Homo sapiens.
 XX
 FN W0200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PP 30-JAN-2001; 2001WO-US00669.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 XX 30-JUN-2000; 2000US-0608408.
 XX 03-AUG-2000; 2000US-0632366.
 XX 21-SEP-2000; 2000US-0234687.
 XX 27-SEP-2000; 2000US-0236359.
 XX 04-OCT-2000; 2000US-0024263.
 PA (MOLB-) MOLECULAR DYNAMICS INC.
 XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 DR Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human fetal liver -
 PT
 XX
 PS Claim 4; SEQ ID NO 16155; 639pp + sequence listing; English.
 CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human fetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 17.1%; Score 241.2; DB 22; Length 436;
 Best Local Similarity 73.2%; Pred. No. 2e-60;
 Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;
 QY 562 TTGTGATCTATGAGACCCCAACAGATATACAGTCCCAACGAGGGAATATGAACTC 621
 DB 1 TTATGTCTATGAGACCCCAACAGATATGAGACCCCACTCCCGATACGAGCC 60
 QY 622 CACGAGAAGATATGAGACCCCAACGAGGATATGAGACCCCACTATGAGATATGAG 681
 DB 61 CACTGCAAGATATGAGACCCCAACGAGGATATGAGACCCCACTATGAGATATGAG 120
 QY 682 CCCCCCTGTGAGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 741
 DB 121 CCTCACCTGTGAGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 180
 QY 742 GAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 801
 DB 181 GAGCCCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 240
 QY 802 ATGTGACCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 861
 DB 241 ATGAGACCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 300
 QY 862 GAAATGAGCCCTACCCCTGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 921
 DB 301 GATACAGAGCTCAGCTGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 356
 QY 922 ACAGATCTATGAGCTCAGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 981
 DB 357 -----CAGGCTCTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 408
 QY 982 ATTACCACTCTCAGAGTAAACCTT 1009
 DB 409 ATTCTTAACCTTAAAGATTAACCTT 436
 RESULT 7
 AAK42003
 ID AAK42003 strand; DNA; 436 BP.
 XX AAK42003;
 AC
 XX 06-NOV-2001 (first entry)
 DT
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 16560.
 XX
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 XX macroarray; cancer; leukemia; lymphoma; myeloma; ss.
 OS Homo sapiens.
 XX

PN WO200157276-A2.
 XX
 XX 09-AUG-2001.
 PD
 XX
 PF 30-JAN-2001; 2001WO-US000668.
 XX
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483900/53.
 DR
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 PS
 PS Example 4; SEQ ID NO: 16560; 658pp + Sequence listing; English.
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 CC
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 17.1%; Score 241.2; DB 22; Length 436;
 Best Local Similarity 73.2%; Pred. No. 2e-60;
 Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;
 QY 562 TTGTGATCTATGAGACCCCAACAGATATACAGTCCCAACGAGGGAATATGAACTC 621
 DB 1 TTATGTCTATGAGACCCCAACAGATATGAGACCCCACTCCCGATACGAGCC 60
 QY 622 CACGAGAAGATATGAGACCCCAACGAGGATATGAGACCCCACTATGAGATATGAG 681
 DB 61 CACTGCAAGATATGAGACCCCAACGAGGATATGAGACCCCACTATGAGATATGAG 120
 QY 682 CCCCCCTGTGAGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 741
 DB 121 CCTCACCTGTGAGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 180
 QY 742 GAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 801
 DB 181 GAGCCCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 240
 QY 802 ATGTGACCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 861
 DB 241 ATGAGACCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 300
 QY 862 GAAATGAGCCCTACCCCTGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 921
 DB 301 GATACAGAGCTCAGCTGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 356
 QY 922 ACAGATCTATGAGCTCAGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 981
 DB 357 -----CAGGCTCTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 408
 QY 982 ATTACCACTCTCAGAGTAAACCTT 1009
 DB 409 ATTCTTAACCTTAAAGATTAACCTT 436
 RESULT 8
 AAI48070

ID AAI48070 standard; DNA; 436 BP.
 XX AAI48070;
 AC
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #16756 used to measure gene expression in human placenta sample.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; as.
 XX
 OS Homo sapiens.
 XX
 PN W0200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLR-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 PD WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID No 16756; 654bp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 XX Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 SQ

Query Match 17.1%; Score 241.2; DB 22; Length 436;
 Best Local Similarity 73.2%; Pred. No. 2e-60;
 Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;

QY 562 TTGTGATCTATGAGACCCCAACAGAGATATACAGTCCACCAAGGGAATATGAAATC 621
 DB 1 TTATTTGCTATGAGACCCCACTGACAGATATGAGGCCCACTCCGGAATACAGCCC 60
 QY 622 CACCAAGAGATATGAGGCCCAACAGGAGATATGAGGCCCACTATGGAATATGAG 681
 DB 61 CACTTGAGAGATATGAGGCCCAACCTTAAAGAAATGAAGGCCCGCTGTGGAATACAGAG 120
 QY 682 CCCGCTGTGAGATATGAGTCCACCTGAGGAGATATGAGTCCCACTTGGGAGATATG 741
 DB 121 CTTCACTGTGAGATATGAGGCCCACTTGTGATATGAGGCCCACTGAGAGATATG 180
 QY 742 GAATCCCACTGGGAGATATGAGGCCCACTTGGGAGATATGAGTCCCACTTGGGAGAT 801
 DB 181 GAGGCCCACTTGAAGATATGAGGCCCACTTGTGATATGAGAACCCCACTCTCCGAT 240
 QY 802 ATGATGCCCACTTGGGAGATATGAGGCCCACTTGAAGATATGAGGCCCACTGAGCTG 861
 DB 241 ATGAGGCCCACTTGTGAGATATGAGGCCCACTTGAAGAAATGAAGGCCCGCTGCGG 300
 QY 862 GAAATGAAGCCCTTACCCCTGCAATATGAGCTCCATCTGCTGGAATATACAGCTGCTCTC 921
 DB 301 GATATGAGAGCTTCACTGCTGAGATCAGAGAGAGGCCCTCAGGAATCTACAGAGCC---- 356

QY 922 ACAGATCTATGAGCTCAGCAGAGACTTCTCCCACTATCATCTTATGAGTCC 981
 DB 357 -----CAGGCTCCGAGAAACAGAGCTTCTTCCCTGCTCTCTTCAAGTCC 408
 QY 982 ATTACCACTTCTCAGAGTTAAACCTT 1009
 DB 409 ATTCTTAACCTTCAAGATGTAAACCTT 436

RESULT 9
 ABS16034
 ID ABS16034 standard; DNA; 436 BP.
 XX
 AC ABS16034;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human genome-derived single exon probe ORF from lung SEQ ID No 16025.
 XX
 KW Human; de; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease; open reading frame; ORF.
 XX
 OS Homo sapiens.
 XX
 PN W0200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLR-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 PD WPI; 2002-114183/15.
 XX
 PT Spatially-addressable set of single exon nucleic acid probes, used to
 XX measure gene expression in human lung samples -
 XX
 PS Claim 4; SEQ ID No 16025; 634bp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included

in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarray having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Rudlak syndrome, sarcoidosis, pulmonary haemoblastosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsagen syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe open reading frame of the invention.

Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_jct_sequences](http://wipo.int/pub/published_jct_sequences).

Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Query Match 17.1%; Score 241.2; DB 24; Length 436;
Best Local Similarity 73.2%; Pred. No. 2e-60;
Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;

```

QY 562 TTGTATCTATGAGACCCCAACGAGATATACATCCCAAGGAGATATGAACTC 621
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 TTATTTCTATGAGGCCCACTGCGAGATATGAGACCCCACTCCGATACGAGCCC 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 622 CACCAAGAGATATGAGGCCCAACGAGGAGATATGAGGCCCACTATGAGATATGAG 681
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 CACCTCAGAGATATGAGGCCCAACGAGGAGATATGAGGCCCACTATGAGATATGAG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 682 CCCCCTGTGAGATATGAGTCCCACTGGGAGATATGAGTCCCACTGGGAGATATG 741
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 CCTCAGCTGTGAGATATGAGGCCCACTTGTGATACGAGGCCCACTGCGAGATATG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 742 GAGTCCCACTGGGAGATATGAGGCCCACTTGGGAGATATGAGTCCCACTGGGAGAT 801
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 GAGCCCCACCTCTAGATATGAGGCCCACTTGTGATATGAGACCCCACTCTGGAT 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 802 ATGGTCCCACTGGGAGATATGAGGCCCACTGAGATATGAGGCCCACTGAGCTG 861
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 ATGAGGCCCACTCTGGAGATATGAGGCCCACTGAGGAAATGAGGCCCGCTGCGG 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 862 GAATGAGCCCTACCCCTGATATGAGCTCATCTGCTGAGAAATACAGCTGCTCTC 921
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 GATACGAGCCCTACCTGCTGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 922 ACGATCTATGAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 981
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 -----CAGGCTCTCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 982 ATTACACCTCTCAGAGTTAACTT 1009
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 409 ATTCTTAACCTTCAAGATGTAACTT 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 10

ID AAL04882 standard; DNA; 7099 BP.

AC AAL04882;

DT 21-NOV-2001 (first entry)

DE Human reproductive system related antigen DNA SEQ ID NO: 7570.

```

XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; de.
XX Homo sapiens.
XX WO20015320-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01339.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 11-JUL-2000; 2000US-0216880.
XX 07-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0232168.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.

```

PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0234984.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246603.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251866.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HOMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM,
 DR WPI, 2001-465570/50.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX
 PS Disclosure; SEQ ID NO 7570; 1297bp + Sequence Listing; English.
 CC
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a genomic sequence encoding a
 CC protein of the invention.
 XX
 SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
 Query Match 15.6%; Score 220; DB 22; Length 7099;
 Best Local Similarity 75.3%; Pred. No. 1.5e-53;
 Matches 274; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
 QY 554 ATATCCAAATTCATATGAGACCCCAACCAAGATATACAGTCCACAGGGGATA 613
 DB 6718 ATCCCAATATTCATATGAGACCCCAACCAAGATATGAGACCCCACTCCGATTA 6777
 QY 614 TGAATCTCACAGAGATATGAGACCCCAACCAAGGGGATATGAGACCCCACTATGG 673
 DB 6778 CGAGCCCACTGAGATATGAGACCCCAACCAAGGGGATATGAGACCCCACTATGG 6837
 QY 674 ATATGAGCCCGCTGAGATATGAGTCCCACTGGGGATATGAGTCCCACTGG 733
 DB 6838 ATACAGAGCTCACTGAGATATGAGACCCCACTGATATGAGAGCCCACTGG 6897
 QY 734 GGGATATGAGTCCCACTGGGGATATGAGACCCCACTGGGGATATGAGTCCCACT 793
 DB 6898 AGGATATGAGCCCACTGAGATATGAGACCCCACTGATATGAGACCCCACT 6957
 QY 794 TGGGATATGAGTCCCACTGGGGATATGAGACCCCACTGAGATATGAGACCC 853
 DB 6958 TCTGGATATGAGACCCCACTGAGATATGAGACCCCACTGAGATATGAGACCC 7017
 QY 854 ACCAGCTGGAATGAGACCCCACTGAGATATGAGTCCCACTGAGATATGAGC 913
 DB 7018 GCGTGGGGATATGAGACCCCACTGAGATATGAGACCCCACTGAGATATGAGC 7077
 QY 914 TGCC 917
 DB 7078 AGCC 7081
 RESULT 11
 ID ABL97776 standard; DNA; 7099 BP.
 XX ABL97776;
 AC
 XX 21-JUN-2002 (first entry)
 DT
 XX Human testicular antigen encoding DNA fragment SEQ ID NO: 2428.
 DB
 XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
 KW reproductive system disorder; urinary system disorder; gene therapy;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disease; infection; cytotoxic; gene; ds.
 XX
 OS Homo sapiens.
 OS
 XX
 FN W0200155317-A2.

XX	MP1, 2001-48332/52.
XX	
PT	Nucleic acids encoding 973 human testicular antigen polypeptides,
PT	useful for preventing, diagnosing and/or treating testicular cancer -
XX	
PS	Disclosure; SEQ ID NO 2428; 766bp; English.
XX	
CC	The present invention provides the protein and coding sequences of 973
CC	human testicular antigens, and fragments of their genomic sequences. The
CC	sequences can be used in the treatment of cardiovascular, urinary system,
CC	reproductive system, immune, respiratory, neurological and
CC	gastrointestinal disorders, infections, and particularly cancer,
CC	especially testicular cancers. The present sequence is a DNA encoding a
CC	protein fragment of the invention.
XX	
SQ	Sequence 7099 BP, 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
XX	
Query Match	15.6%; Score 220; DB 23; Length 7099;
Best Local Similarity	75.3%; Pred. 1.5e-53;
Matches 274; Conservative	0; Mismatches 90; Indels 0; Gaps 0;
OY	554 ATATCAATTGTGATCTATGACCCCAACACAGATATACAGTCCACAGGGGAATA 613
DB	6718 ATCCCACTTATATGCTATATGAGCCCACTGACAGATATGAGCCCACTCCGAGTA 6777
OY	614 TGAATCTCACAGAGATATGAGCCCAACAGGGGATATGAGCCCACTATGGG 673
DB	6778 CGAGACCCCACTGACAGATATGAGCCCAACCTGAGAAATGAAGCCCGCTGTGGG 6837
OY	674 ATATGAGACCCCGCTGTGGGATATGAGAGTCCACTGGGGATATGAGTCCCACTGG 733
DB	6838 ATACAGAGCTCACTGTGCATATGAGAGCCCACTCTTGATATGAGAGCCCACTGCG 6897
OY	734 GGGATATGAGAGTCCCACTGGGGATATGAGAGCCCACTGGGGATATGAGTCCCACT 793
DB	6898 AGGATATGAGAGCCCACTCTAGATATGAGAGCCCACTCTGTGATATGAGAGCCCACT 6957
OY	794 TGGGGATATGAGTCCCACTGGGGATATGAGAGCCCACTGACAGATATGAGAGCC 853
DB	6958 TCTCGAATATGAGAGCCCACTCTGAGATATGAGAGCCCACTGACAGAAATGAAGCC 7017
OY	854 ACCAGCTGGAATGAAGCCCTACCCCTGACATATGAGAGTCAATGCTGAGAAATACAG 913
DB	7018 GCTTCGGAGATACAGAGCTCACTGTGATGATACAGAGCAAGCTCAAGAAATCAAGC 7077
OY	914 TGCC 917
DB	7078 AGCC 7081
XX	
RESULT 12	
AL01355	
ID	AL01355 standard; cDNA; 321 BP.
XX	
AC	AL01355;
XX	
DT	21-NOV-2001 (first entry)
XX	
DE	Human reproductive system related antigen cDNA SEQ ID NO: 1356.
XX	
KM	Human; reproductive system related antigen; reproductive system disorder;
XX	cancer; gene therapy; ss.
OS	Homo sapiens.
XX	
FN	WO200155320-A2.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001MO-US01339.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.

PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-465570/50.
 DR P-PSDB; AAM95385.
 XX
 XX Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX
 PS Claim 1; SEQ ID NO 1356; 1297bp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a coding sequence of the
 CC invention.
 XX
 SQ Sequence 321 BP; 83 A; 75 C; 70 G; 89 T; 4 other;
 Query Match 13.8%; Score 195.4; DB 22; Length 321;
 Best Local Similarity 77.4%; Pred. No. Se-47; Mismatches 69; Indels 1; Gaps 1;
 Matches 246; Conservative 2;
 QY 64 GCCGTGAGGAGCCCTCATCCCTGCGGAAAGTGTGTAAGCAGTGTGAGATGCG 123
 DB 1 GCCGCCGCGGAGCCCTCATCCCTGCGGAAAGTGTGTAAGCAGTGTGAGATGCG 60
 QY 124 ACCCTGCTCTCTACGAAACCAAGTGAATCTATCTTTAAGCACAAGAAAGAA 183
 DB 61 AGCT 120
 QY 184 CGTTGTTCTCACTTATACCGGGGCTTCTGACTTCACTTATGCAATGACCCCA 243
 DB 121 CATGTTCTCTCACTTATACCGGGGCTTCTGACTTCACTTATGCAATGACCCCA 180
 QY 244 TGCCTTCTTATGATGCGTGTGCTGATGAGTCACTGACCAATGACCAATTT 303
 DB 181 TGTGCTTTTATGATGCGTGTGCTGATGAGTCACTGACCAATGACCAATTT 240
 QY 304 TTGCCCCCACTTATTAAGAACCAATTCAGGACCTCCAGTGTGCTGGG-AGGA 362
 DB 241 TTGCTGCAAACTTCAATTAAGGAACTATTCAGGACCTCCAGTGTGCTGGG-AGGA 300
 QY 363 CAAGCTTTTAAAGTTA 380
 DB 301 CAAGCTTTTAAAGTTA 318
 RESULT 13
 ABL96808
 ID ABL96808 standard; cDNA, 321 BP.
 XX
 AC ABL96808;
 XX
 DT 21-JUN-2002 (first entry)
 XX
 XX Human testicular antigen encoding cDNA SEQ ID NO: 476.
 DE
 XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
 KW reproductive system disorder; urinary system disorder; gene therapy;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disease; infection; cytostatic; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX MO200155317-A2.
 PN
 PD 02-AUG-2001.
 XX
 XX 17-JAN-2001; 2001WO-US01329.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.

```
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225271.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251889.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(PHMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483232/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
PT useful for preventing, diagnosing and/or treating testicular cancer -
XX
XX Claim 1; SEQ ID NO 476, 766pp; English.
XX
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a cDNA of the
CC invention.
XX
XX Sequence 321 BP; 83 A; 75 C; 70 G; 89 T; 4 other;
```

Query Match 13.8%; Score 195.4; DB 23; Length 321;
Best Local Similarity 77.4%; Pred. No. 5e-47;
Matches 246; Conservative 2; Mismatches 69; Indels 1; Gaps 1;

QY 64 GCGGCTGTGGGCGCTCATCCCTCTGGCGAAAGTGTGAAAGCATGTAGATGTGG 123
DB 1 GCGGCGCGGAGCGCTCATCCCTTACCGTGAAGTCTGTGAAGCGGTCTCGAATGTGG 60
QY 124 AACTGCTTCCACAGAAACAGTGAATCCATCTCTTTATGGCAAAAGAAAGAA 183
DB 61 AGCTCTCTCCACAGCAATCAAGAGCTCAAAATCTTATGTGTAGAAAGCAAGAA 120
QY 184 CGTGTGTCTCATCTTATACCGGCTGTCTGTGACTTCACTTAAAGTCAATGACCCCA 243
DB 121 CATGTGTCTCATCTTATACCGGCTGTCTGTGACTTCACTTAAAGTCAATGACCCCA 180
QY 244 TCGTTCCTTTATGATGCGGCTGTGGCTGATGATGATGATGATGATGATGATGAT 303
DB 181 TGTGTCTTTATGATGCGGCTGTGGCTGATGATGATGATGATGATGATGATGAT 240
QY 304 TTGCCCCCACTCATTAAGAAACCATTCAGGCACTCCAGTGTGGCTGGG-AAAGA 362
DB 241 TTGCTGCAAACTTCATTAAGGAACTATTCAGGCACTTCATTTGGGCTGGGAAAGA 300
QY 363 CAAGCTGTTTTTAAGTTA 380
DB 301 CAAGCTACTTTTAANTTA 318

RESULT 14
ABST0430
ID ABST0430 standard; cDNA; 1885 BP.
XX AC ABST0430;
XX DT 27-NOV-2002 (first entry)
XX DE Human bone remodelling gene #87.
XX KW Bone remodelling; osteoporosis; human; gene; 88.
XX OS Homo sapiens.
XX PN US6426186-B1.
XX PD 30-JUL-2002.
XX PF 18-JAN-2000; 2000US-0484970.
XX PR 18-JAN-2000; 2000US-0484970.
XX PA (INCY-) INCYTE GENOMICS INC.
XX PI Jones KA, Volkmut W, Walker MG;
XX DR WPI; 2002-673014/72.
XX PT A combination of polynucleotides which are co-expressed with genes
XX known to be involved in bone remodeling and osteoporosis are useful in
XX an array for the diagnosis of bone remodeling and osteoporosis
XX associated disorders -
XX PS Claim 1; Column 247-250; 206pp; English.
XX CC The invention relates to a combination comprising a number of
XX co-expressed with genes known to be involved in bone remodeling and
XX osteoporosis. The invention is used to diagnose disorders associated
XX with bone remodeling or osteoporosis. ABST0344-ABST0512 represent
XX human bone remodelling genes of the invention.
XX SQ Sequence 1885 BP; 384 A; 610 C; 476 G; 415 T; 0 other;

Query Match 8.6%; Score 122; DB 24; Length 1885;
Best Local Similarity 59.4%; Pred. No. 5e-25;
Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;

QY 36 ATGCAATGAAACCAAGGCAACGAGAGCGGTGTGGGCGCTCATCCCTCTGGCGAA 95
DB 61 ATGGGCTTCACAAAGAACTCTCGAGGCGCGGAGATGATGATGATGATGATGATGAT 117
QY 96 AGTGTCTTGAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 155
DB 118 AGCATCTTAATGTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 177
QY 156 TATCTTTTAAATGCAAAAGAAAGAAAGTGTGTTCTCACTTCAATCCGGGTGTCTTC 215
DB 178 GAAGCTTCAAAAGGCAAGAAAGGCACTGTCTTACCTTACCCCTTACGGGTGTCTTC 237
QY 216 GTGACTTCACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 275
DB 238 CTGTCT 291
QY 276 AGTACTGCAACATTTGAACCAACCAATTTTGGCCCCCACTAATTAAAGAAACATTGAG 335
DB 232 AAGACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 351
QY 336 GCACTCCAGGCTGTGGCTGGGAAAGCAAGCTGTTTTTAAGTTATCTTCAAGAAAGA 395
DB 352 GCGGAAGGGGAGGCTGTGGGAAAGCTGTCTTCTCAAGTTGACTTCAAGGCAAGG 411
QY 396 GGTGCAATGCAATTTGCCCACTGATGATGATGATGATGATGATGATGATGATGAT 451
DB 412 GCGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 467

RESULT 15
AAF22363
ID AAF22363 standard; cDNA; 1915 BP.
XX AC AAF22363;
XX DT 26-MAR-2001 (first entry)
XX DE Human secreted protein gene 48 SEQ ID NO:58.
XX KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytoskeletal; cardiac; vasotropic;
XX cerebroprotective; noctropic; neuroprotective; antibacterial; vitruide;
XX fungicide; ophthalmological; vulnary; gene therapy; neoplasm;
XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
XX cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
XX cerebral ischaemia; angiodenesis; nervous system disorder; infection;
XX Alzheimer's disease; ocular disorder; corneal infection; wound healing;
XX skin aging; food additive; preservative; 88.
XX OS Homo sapiens.
XX PN WO200061748-A1.
XX PD 19-OCT-2000.
XX PF 06-APR-2000; 2000WO-US08982.
XX PR 09-APR-1999; 99US-0128696.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben SM, Komatsoulis G;
XX DR WPI; 2000-638566/61.
XX DR P-PSDB; AAB63096.
XX PT New nucleic acid molecules encoding 48 human secreted proteins for
XX diagnosing, preventing, treating or ameliorating medical conditions and

PT used as food additives or preservatives -
XX
PS Claim 1; Page 429-430; 480pp; English.

AAFP22316 to AAF22263 encode the human secreted proteins given in AAB63049 to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins and polypeptides homologous to them. Human secreted proteins have activities based on the tissues and cells the genes are expressed in. Examples of activities include: immunosuppressive; antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial; vitnucide; fungicide; ophthalmological; and vulnereary. The polynucleotides and proteins can be used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. AAFP22307 to AAF22215 and AAB63048 represent sequences used in the exemplification of the present invention.

SQ Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;

Query Match	8.6%	Score 122;	DB 21;	Length 1915;
Best Local Similarity	59.4%;	Pred. No. 5e-25;		
Matches 247; Conservative	0;	Mismatches 160;	Indels 9;	Gaps 2

Oy 36 ATGGCAGTGAACCAAGAGCCACACCGAGAGCCGTGCGGGGCCCTCATCCCTCTGGCGAA 95
 Db 59 ATGGCGCTCAACAAATATCACTCGAGGGCGCGAGTGTATGTCATAAAGC---CGAG 115
 Oy 96 AGTGTCTTGAAAGCAGTGTAGGAGTGTGACCTCTGCTTCTACAGAAACCAAGTGAATCC 155
 Db 116 AGCATCTTAATGTCTTATGATCACTGGAACCTCAATTCAGTAACATGAAGAAGCGGCCA 175
 Oy 156 TATCTCTTAAATGCGACAAAGAAAGGAACGTTGTTCTCACTTCATACCGGGTGGTCTTC 215
 Db 176 GAAGCTTCAAAAGGAGCCAAAGAAAGGACCTGTCTACCTTAACCCCTTAACGGGGTCATCTTT 235
 Oy 216 GTGACCTTCAACCTTAGTCATATACCCCATGCTTTCTTTATATGATGCGTTTGGCCCTGATG 275
 Db 236 CTGTCT-----CAAGGGCAGAGATGCCATAGCAAGTCTTCAATGATGCAATTTATATTCATG 289
 Oy 276 AGTGACTGCACCATTTGAACAAACCAATTTTGTGCCCCCACTACATTAAGAAACCATTCAG 335
 Db 290 AAAGACTGTGAATATCAAGCAGGCCCGATATTTGTGTGCAAACTACATCAAGGAAACAGTGAAG 349
 Oy 336 GCAGCTCCAGGTGTGTGCTGGGAAAGGACAAAGCTGTTTAAAGTTATCTTCAGAAAGA 395
 Db 350 GCGGAAGCGGAGGTGTGCTGGGAAAGGCTCTGCTCTCTCAAGATTGACTTTCACGCGAGGG 409
 Oy 396 GGTGCAATCGAATTTTGGCCAACTGATGTTAAAGCTGCTCTGTGCTGTCCAGAG 451
 Db 410 GCGGCATTTAGTTTGGACAGCGAGATGCTTCAGTGTGCATCTCAAGCTTCCAGAG 465

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:55:15 (Search time 99.5467 Seconds)

(Without alignments)
6265.145 Million cell updates/sec

Title: US-09-864-291-4

Sequence: 1413
1 cgcacgagagggcgagga.....gcataaaaaaaaaaaaaa 1413

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

Issued Patents NA: *
1: /cgn2_6/ptodata/2/1na/5A COMB.seq.*
2: /cgn2_6/ptodata/2/1na/5B COMB.seq.*
3: /cgn2_6/ptodata/2/1na/6A COMB.seq.*
4: /cgn2_6/ptodata/2/1na/6B COMB.seq.*
5: /cgn2_6/ptodata/2/1na/PTCUS COMB.seq.*
6: /cgn2_6/ptodata/2/1na/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122	8.6	1885	4 US-09-484-970B-87	Sequence 87, Appl
2	64.8	4.6	981	2 US-08-841-349-15	Sequence 15, Appl
3	59.8	4.2	3489	2 US-08-728-123A-1	Sequence 1, Appl
4	59.8	4.2	3489	4 US-09-298-568-1	Sequence 1, Appl
5	59.8	4.2	3489	4 US-09-410-399-1	Sequence 1, Appl
6	59.8	4.2	32207	2 US-08-770-379-20	Sequence 20, Appl
7	59.8	4.2	32207	3 US-08-757-669A-20	Sequence 20, Appl
8	59.8	4.2	32207	4 US-09-230-371A-20	Sequence 20, Appl
9	54.8	3.9	3833	1 US-08-917-320-18	Sequence 18, Appl
10	54.8	3.9	3833	5 PCT-US95-04611A-18	Sequence 18, Appl
11	54.8	3.9	5931	3 US-08-783-774-1	Sequence 1, Appl
12	54.8	3.9	5931	4 US-09-556-706B-1	Sequence 1, Appl
13	53.4	3.8	243	1 US-08-182-175A-56	Sequence 56, Appl
14	53.4	3.8	243	1 US-08-474-633A-74	Sequence 74, Appl
15	53.4	3.8	243	4 US-08-823-771-74	Sequence 74, Appl
16	53.4	3.8	243	5 PCT-US92-06412-56	Sequence 56, Appl
17	53	3.8	2144	3 US-08-834-306-15	Sequence 15, Appl
18	53	3.8	2144	4 US-09-256-976-15	Sequence 15, Appl
19	53	3.8	2144	3 US-08-993-674A-15	Sequence 15, Appl
20	52.4	3.7	2793	1 US-08-209-747-1	Sequence 1, Appl
21	52.4	3.7	2793	1 US-08-458-398-1	Sequence 1, Appl
22	51.2	3.6	3794	4 US-09-192-434-1	Sequence 1, Appl
23	49.8	3.5	1235	2 US-08-557-309B-53	Sequence 53, Appl
24	49.6	3.5	187	1 US-08-182-175A-90	Sequence 90, Appl
25	49.6	3.5	187	1 US-08-474-633A-78	Sequence 78, Appl
26	49.6	3.5	187	4 US-08-823-771-78	Sequence 78, Appl
27	49.6	3.5	187	5 PCT-US92-06412-90	Sequence 90, Appl

28	49.4	3.5	8310	3 US-08-870-126-11	Sequence 11, Appl
29	49.4	3.5	8310	4 US-09-445-247-11	Sequence 11, Appl
30	49.4	3.5	14985	1 US-08-652-972A-6	Sequence 6, Appl
31	49.4	3.5	14985	5 PCT-US96-06231A-6	Sequence 6, Appl
32	49.2	3.5	2188	1 US-07-865-662B-10	Sequence 10, Appl
33	49.2	3.5	2188	3 US-08-374-219B-10	Sequence 10, Appl
34	48.8	3.5	5661	4 US-08-938-105-2	Sequence 2, Appl
35	48.6	3.4	16442	3 US-08-781-891-208	Sequence 208, App
36	48.6	3.4	16442	4 US-09-618-166-209	Sequence 209, App
37	48.6	3.4	51259	3 US-08-781-891-209	Sequence 209, App
38	48.6	3.4	51259	4 US-09-618-166-209	Sequence 209, App
39	47.4	3.4	471	4 US-09-370-838-278	Sequence 278, App
40	47.4	3.4	977	6 5215895-2	Patent No. 5215895
41	46.2	3.3	456	2 US-08-557-309B-16	Sequence 16, Appl
42	46.2	3.3	456	3 US-08-834-306-16	Sequence 16, Appl
43	46.2	3.3	456	3 US-08-993-674A-16	Sequence 16, Appl
44	46.2	3.3	456	4 US-09-256-976-16	Sequence 16, Appl
45	45.8	3.2	977	1 US-08-017-522A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-484-970B-87
Sequence 87, Application US/09484970B
Patent No. 6426186
GENERAL INFORMATION:
APPLICANT: Jones, Karen A.
APPLICANT: Volkmut, Wayne
APPLICANT: Walker, Michael G.
TITLE OF INVENTION: BOMB REMODELING GENES
FILE REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 1885
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 232968.10CB1
US-09-484-970B-87

Query Match	8.6%; Score 122; DB 4; Length 1885;
Best Local Similarity	59.4%; Pred. No. 1.7e-27;
Matches	247; Conservative 160; Indels 9; Gaps 2;
QY	36 ATGGCAGTGAACGAGCAACCGAGCCGCTGCGGACCTCATCCCTCTGCGAA 95
DB	61 ATGGCCTCAACGAGATCACTCGAGCGCGAGATATGCTCATTAACAC---CGAG 117
QY	96 AGTGTCTGAAGAGTGAAGATGAGATCTCTGCTCTCTACAGAAACAGTGAATCC 155
DB	118 AGATCTATATGCTTATATATCACTGGAATCACTTAATGATGAAAGCGGCA 177
QY	156 TATCTCTTAAATGCAAGAAAGAAAGATGTTTCTCATCTTATACCGGATGCTTC 215
DB	178 GAAGCTTCAAGAGGACCAAGAAAGGACATGCTTACCTTACCCCTTACCGGATCTTT 237
QY	216 GTGACTTCACTTACTGATGATGCCCATGCTTTCTTATATGATGCGTTTGCGTGAATG 275
DB	238 CTGTCTC-----CAAGGCAAGATGCAAGCATGCTTCTCATATATGTCATTTATCTCATG 291
QY	276 AGTGAATGACCATTTGAACCAATTTTGGCCCACTTACATTAAGAAACCATTCAG 335
DB	292 AAGACCTGAGATGATCAAGAGCCCGTATTTGTCGAACATCAATCAAGGAACAGTGAAG 351
QY	336 GCAGCTCCAGTGGTGGCTGGGAGGACAAAGCTGTTTAAAGTATCTTCCAGAAAGGA 395
DB	352 GCGGAACGCGGAGGTGGCTGGAGGCTGCTTCTTACAGTTGACTTTCACGAGGAGG 411


```

: TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
: FILE REFERENCE: 16412-10001R
: CURRENT APPLICATION NUMBER: US/09/298,568
: CURRENT FILING DATE: 1999-04-21
: EARLIER APPLICATION NUMBER: US 60/409,422
: EARLIER FILING DATE: 1998-11-19
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3489
: TYPE: DNA
: ORGANISM: Kaposi's sarcoma-associated herpesvirus
: US-09-298-568-1

```

Query Match	4.2%;	Score 59.8;	DB 4;	Length 3489;
Best Local Similarity	46.9%;	Pred. No. 4.9e-08;		
Matches 187; Conservative	0;	Mismatches 212;	Indels 0;	Gaps 0;

QY	514	CAGGGGGCTGCAAGTGTGCTCTCCACAGAACCTTTGTCGAGCATATCAATTGGATCTATG	573
Db	2185	CAGCAGCAGCAGAGTATGACGACGACGACGATGAGCAGCAGCAGCAGAGATGAGCAGCAG	2244
QY	574	GAACCCCAACCAACAGATATACGTCCACCAAGGGGAAATATGAACTCCACCAAGAGAT	633
Db	2245	CAGCAGAGATGAACAGGACAGCAGGAGGACGAGCAGCAGAGAGACAGCAGGACGAGAG	2304
QY	634	ATGCAAGCCCAACAGGGGGATATGAGGCCCACTATATGGATATATGAGACCCGCTGTGG	693
Db	2305	TTAGAGGAGCAGAGACAGAGAGTTAAGAGATCAGGAGCAGAGATTATGAGAGCAGAGCAG	2364
QY	694	GATATGAGATCCCACTGGGGATATGAGATCCCACTGGGGGATATGAGATCCCACTG	753
Db	2365	GAGTTAGAGAGCAGAGCAGAGAGTTAAGAGAGCAGAGCAGAGATTTAGAGAGCAGAG	2422
QY	754	GGGGATATGAGACCCCACTGGGGGATATGAGTCCCACTGGGGGATATGATGCCCA	813
Db	2425	CAGAGTTAAGAGAGCAGAGCAGAGAGTTAAGAGAGCAGAGCAGAGAGTTAAGAGCAG	2484
QY	814	CTGGGGGATATGAGACCCCACTGCAAGATATGAGACCCCACTGAGTGAATGAAGCC	873
Db	2485	GAGCAGGAGTTAAGAGAGCAGAGCAGAGGTTAAGAGAGCAGAGGTGAAGAGCAGAG	2544
QY	874	TACCCCTGCAATGAGCTCATCTGTGAGAAATACAG	912
Db	2545	CAGAGGTGAGAGACAAAGCAGAGCAGAGAGAGCAG	2583

```

RESULT 5
US-09-410-399-1
: Sequence 1, Application US/09410399
: Patent No. 6482587
: GENERAL INFORMATION:
: APPLICANT: Robertson, Erle S.
: APPLICANT: Coteer, Murray A.
: TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
: TITLE OF INVENTION: to Genomic Host DNA
: FILE REFERENCE: UM-03778
: CURRENT APPLICATION NUMBER: US/09/410,399
: CURRENT FILING DATE: 1999-10-01
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3489
: TYPE: DNA
: ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

```

Query Match	4.2%;	Score 59.8;	DB 4;	Length 3489;
Best Local Similarity	46.9%;	Pred. No. 4.9e-08;		
Matches 187; Conservative	0;	Mismatches 212;	Indels 0;	Gaps 0;

514 CAGGGGCTGCAGTGTCTCTCTCAGACACACCTTGTCCAGCATTCCATTGTCATCTATG 5/3

[illegible]

RESULT 6
US-08-770-379-20/c
! Sequence 20, Application US/08770379

1 Patent NO. 5849564
2
3 GENERAL INFORMATION:
4 APPLICANT: Chang, Yuan
5 APPLICANT: Bohenzky, Roy A.
6 APPLICANT: Russo, James J.
7 APPLICANT: Edelman, Isidore S.
8 APPLICANT: Moore, Patrick S.
9 TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
10 TITL OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
11 NUMBER OF SEQUENCES: 20
12 CORRESPONDENCE ADDRESSES:
13 ADDRESSEE: Cooper & Dunham LLP
14 STREET: 1185 Avenue of the Americas
15 CITY: New York
16 STATE: New York
17 COUNTRY: U.S.A.
18
19 ZIP: 10036
20
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23 OPERATING SYSTEM: IBM PC compatible
24 SOFTWARE: PatentIn release #1.0, Version #1.30
25 CURRENT APPLICATION DATA:
26 APPLICATION NUMBER: US/08/770,379
27 FILING DATE:
28

```

1 CLASSIFICATION: 435
2
3 ATTORNEY/AGENT INFORMATION:
4
5   NAME: White, John P.
6
7   REGISTRATION NUMBER: 28,674
8
9   REFERENCE/DOCKET NUMBER: 5
10
11 TELECOMMUNICATION INFORMATION:
12
13   TELEPHONE: (212) 278-0400
14
15   TELEFAX: (212) 391-0525
16
17 INFORMATION FOR SEQ ID NO: 20
18
19   SEQUENCE CHARACTERISTICS:
20
21     LENGTH: 32207 base pairs
22
23     TYPE: nucleic acid
24
25     STRANDEDNESS: double
26
27     TOPOLOGY: linear
28
29     MOLECULE TYPE: DNA (genomic)
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
101
```

Query Match 4.2%; Score 59.8; DB 2; Length 32207;

Best Local Similarity 46.9%; Pred. No. 1.9e-07;
Matches 187, Conservative 0; Mismatches 212, Indels 0, Gaps 0;
Db 514 CAGGGGCTGAGTGTCTCTCTCAAGACACCTTGTCCAGATATCCAAATTGTATCTATG 573
19812 CAGCAGCAGAGATGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAG 19753
Qy 574 GACCCCAACCAACAGATATACGTCCAAACAGGGGAAATATGAACTCCACCAAGAGAT 633
Db 19752 CAGCAGAGATGAAACAGAGCAGCAGAGAGCAGAGCAGACAGCAGAGAGCAGAG 19693
Qy 634 ATGAGCCCAACCAAGAGATATACGTCCAAACAGGGGAAATATGAACTCCACCAAGAGAT 633
Db 19692 TTAGAGAGAGCAGAGCAGAGATTAAGAGATCAGAGAGAGATTAAGAGAGCAGAG 19633
Qy 694 GATATGAGTCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 753
Db 19632 GAGTTAGAGAGCAGAGCAGAGATTAAGAGAGCAGAGAGAGATTAAGAGAGCAGAG 19573
Qy 754 GGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 813
Db 19572 CAGAGATTAGAGAGCAGAGCAGAGATTAAGAGAGCAGAGAGAGATTAAGAGAGCAG 19513
Qy 814 CTGGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 873
Db 19512 GAGCAGAGATTAGAGAGCAGAGCAGAGATTAAGAGAGCAGAGAGATTAAGAGAGCAG 19453
Qy 874 TACCCCTGATATGAGCTTCATCTGCTGGAATATAG 912
Db 19452 CAGAGATTAGAGAGCAGAGCAGAGCAGAGCAGAGAGAGCAG 19414

RESULT 7
US-08-757-669A-20/c
Sequence 20, Application US/08757669A
Patent No. 6183751
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNICORE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match
Best Local Similarity 46.9%; Pred. No. 1.9e-07;
Matches 187, Conservative 0; Mismatches 212, Indels 0, Gaps 0;

Qy 514 CAGGGGCTGAGTGTCTCTCTCAAGACACCTTGTCCAGATATCCAAATTGTATCTATG 573
Db 19812 CAGCAGCAGAGATGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAG 19753
Qy 574 GACCCCAACCAACAGATATACGTCCAAACAGGGGAAATATGAACTCCACCAAGAGAT 633
Db 19752 CAGCAGAGATGAAACAGAGCAGAGAGAGCAGAGCAGAGCAGAGAGAGCAGAG 19693
Qy 634 ATGAGCCCAACCAAGAGATATACGTCCCACTTATGAGTCCCACTGAGGGGATATGAGTCCCACTG 693
Db 19692 TTAGAGAGAGCAGAGCAGAGATTAAGAGATCAGAGAGAGATTAAGAGAGCAGAG 19633
Qy 694 GATATGAGTCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 753
Db 19632 GAGTTAGAGAGCAGAGCAGAGATTAAGAGAGCAGAGAGAGATTAAGAGAGCAGAG 19573
Qy 754 GGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 813
Db 19572 CAGAGATTAGAGAGCAGAGCAGAGATTAAGAGAGCAGAGAGAGATTAAGAGAGCAG 19513
Qy 814 CTGGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 873
Db 19512 GAGCAGAGATTAGAGAGCAGAGCAGAGATTAAGAGAGCAGAGAGATTAAGAGAGCAG 19453
Qy 874 TACCCCTGATATGAGCTTCATCTGCTGGAATATAG 912
Db 19452 CAGAGATTAGAGAGCAGAGCAGAGCAGAGCAGAGAGAGCAG 19414

RESULT 8
US-09-230-371A-20/c
Sequence 20, Application US/09230371A
Patent No. 6348586
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNICORE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
TITLE OF INVENTION: USES THEREOF
FILE REFERENCE: 45185-G-PCT-US
CURRENT FILING DATE: US/09/230,371A
CURRENT FILING DATE: 1999-11-17
PRIOR APPLICATION NUMBER: PCT/US97/13346
PRIOR FILING DATE: 1997-07-22
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 20
LENGTH: 32207
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

Query Match
Best Local Similarity 46.9%; Pred. No. 1.9e-07;
Matches 187, Conservative 0; Mismatches 212, Indels 0, Gaps 0;

Qy 514 CAGGGGCTGAGTGTCTCTCTCAAGACACCTTGTCCAGATATCCAAATTGTATCTATG 573
Db 19812 CAGCAGCAGAGATGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAG 19753
Qy 574 GACCCCAACCAACAGATATACGTCCAAACAGGGGAAATATGAACTCCACCAAGAGAT 633
Db 19752 CAGCAGAGATGAAACAGAGCAGAGAGAGCAGAGCAGAGCAGAGAGAGCAGAG 19693

QY 634 ATGAGGCCCAACGAGGGGATATGAGCCCACTATGAGTATGAGCCCCCTGTG 693
DB 19692 TTATAGAGGACGAGGAGTTAGAGATGAGAGCAGAGGATTAGAGAGCAGAGCAG 19633
QY 694 GATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 753
DB 19632 GAGTTAGAGGAGCAGAGGAGTTAGAGAGCAGAGGAGTTAGAGAGCAGAG 19573
QY 754 GGGATATGAGCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 813
DB 19572 CAGAGTTAGAGGAGCAGAGGAGTTAGAGAGCAGAGGAGTTAGAGAGCAG 19513
QY 814 CTGGGGATATGAGCCCACTGAGGATATGAGCCCACTGAGGATATGAGCCCACTG 873
DB 19512 GAGAGGAGTTAGAGGAGCAGAGGAGTTAGAGAGCAGAGGAGTTAGAGAGCAG 19453
QY 874 TACCCCTGATGAGCTCATCTGCTGGAATAC 912
DB 19452 CAGAGGTTGAGAGCAGAGCAGAGCAGAGCAGAGCAGAGCAG 19414

RESULT 9

US-08-917-320-18

Sequence 18, Application US/08917320

Patent No. 5824508

GENERAL INFORMATION:

APPLICANT: Spaete, Richard and Jackman, Winthrop, T.

TITLE OF INVENTION: No. 5824508 Splicing Variants of gp350/220

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESSES:

STREET: 5 Palo Alto Square

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/917,320

CLASSIFICATION: 435

FILING DATE: 25-AUG-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/229,291

FILING DATE: April 18, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Luann Gaert

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR-003/00US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-843-5163

TELEFAX: 415-857-0663

TELEX: 380816 COOLEYPA

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 3833 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1014..3734

US-08-917-320-18

Query Match 3.9%; Score 54.8; DB 1; Length 3833;
Best Local Similarity 48.7%; Pred. No. 1.8e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 551 AGCATATCCATTGTGATCTATGACCCCAACCAAGGATATACATCCACCAAGGGA 610
DB 2532 ACCAGCTCCACTCACCAGTACTACCCCAACCCCAATGCCACAGCCCACTCCAGCA 2591
QY 611 ATATGAACTCCACGAAAGATATGAGCCCAACGAGGAGATATGAGCCCACTAT 670
DB 2592 GTGACTTACCCCAACCCCAATGCCACCAACCCCACTCCAGTACTACCCCAACCCCA 2651
QY 671 GGGATATGAGCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 730
DB 2652 AATGCCACGAGCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 711
QY 731 TGGGGATATGAGTCCCACTGAGGATATGAGCCCACTGAGGATATGAGTCCCACTG 790
DB 2712 CCAATGCCACGAGCCCACTGAGGATATGAGCCCACTGAGGATATGAGTCCCACTG 2771
QY 791 ACCTGGGATATGAGTCCCACTGAGGATATGAGCCCACTGAGGATATGAGTCCCACTG 850
DB 2772 ACCCAATGCCACGAGCCCACTGAGGATATGAGCCCACTGAGGATATGAGTCCCACTG 2831
QY 851 CCCACC 856
DB 2832 CCAACC 2837

RESULT 10

PCT-US95-04611A-18

Sequence 18, Application PC/TUS9504611A

GENERAL INFORMATION:

APPLICANT: Spaete, Richard and Jackman, Winthrop, T.

TITLE OF INVENTION: Non Splicing Variants of gp350/220

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESSES:

STREET: 5 Palo Alto Square

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04611A

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/229,291

FILING DATE: April 18, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Luann Gaert

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: AVIR-003/00US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-843-5163

TELEFAX: 415-857-0663

TELEX: 380816 COOLEYPA

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 3833 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1014..3734

PCT-US95-04611A-18

Query Match 3.9%; Score 54.8; DB 5; Length 3833;

Db 2712 CCAATGCGACAGAGCCCACTTGGGAAAAAAGAGCCCACTGAGAGTGAATACCCA 2771
QY 791 ACCCTGGGGGATATGATGAGGAGTATGAGAGCCCACTGAGAGTATGAGAC 850
Db 2772 ACCCGAATGCGACAGAGCCCACTTGGGAAAAAAGAGCCCACTGAGAGTGAATAC 2831
QY 851 CCCAGC 856
Db 2832 CCAAGC 2837

RESULT 13

US-08-182-175A-56
; Sequence 56, Application US/08182175A
; Patent No. 5559223
; GENERAL INFORMATION:
; APPLICANT: Saverio Carl Falco
; APPLICANT: Sharon J. Keeler
; APPLICANT: Janet A. Rice
; TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
; NUMBER OF SEQUENCES: 113
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. du Pont de Nemours and Company
; STREET: 1007 Market Street
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh System, 6.0
; SOFTWARE: Microsoft Word, 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/182,175A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/743,006
; FILING DATE: 9 August 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Linda Akemethy Floyd
; REGISTRATION NUMBER: 33,692
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (302) 992-4929
; TELEFAX: (302) 892-7949
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: E. coli
; CELL TYPE: DH5 alpha
; IMMEDIATE SOURCE:
; CLONE: 2-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..235
; OTHER INFORMATION: /function= "synthetic storage protein"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "asg"
; OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-182-175A-56

Query Match 3.8%; Score 53.4; DB 1; Length 243;
Best Local Similarity 53.0%; Pred. No. 9.3e-07;
Matches 114; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 634 ATGAGCCCAACCAAGGGGATATGAGAGCCCACTTATGGATATGAGAGCCCGCTGTGG 693
Db 2 ATGAGAGAGAGAGCTGAAGAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61
QY 694 GATATGAGATCCCACTGGGGGATATGAGAGTCCCACTGGGGGATATGAGAGTCCCACTG 753
Db 62 GCATGAGAGAGAGAGCTGAAGAGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTG 121
QY 754 GGGATATGAGAGCCCACTGGGGGATATGAGAGTCCCACTGGGGGATATGAGAGTCCCACTG 813
Db 122 AAGCGATGAA 181
QY 814 CTGGGGGATATGAGAGCCCACTGAGAGATGGA 848
Db 182 CTCAAAAAGATGAG 216

RESULT 14

US-08-474-633A-74
; Sequence 74, Application US/08474633A
; Patent No. 5773691
; GENERAL INFORMATION:
; APPLICANT: E. I. DU PONT DE NEMOURS AND
; APPLICANT: COMPANY
; TITLE OF INVENTION: CHIMERIC GENES AND
; TITLE OF INVENTION: METHODS FOR INCREASING
; TITLE OF INVENTION: INCREASING THE LYSINE
; TITLE OF INVENTION: AND THREONINE CONTENT
; TITLE OF INVENTION: OF THE SEEDS OF PLANTS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS
; ADDRESS: AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: U.S.A.
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD VERSION 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,633A
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: BARBARA C. SIEGEL
; REGISTRATION NUMBER: 30,684
; REFERENCE/DOCKET NUMBER: BB-1037-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-4931
; TELEFAX: 302-773-0164
; TELEX: 835420
; INFORMATION FOR SEQ ID NO: 74:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; STRAIN: E. coli
; CELL TYPE: DH5 alpha
; IMMEDIATE SOURCE:
; CLONE: 2-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..235
; OTHER INFORMATION: /function= "synthetic
; OTHER INFORMATION: storage protein"

OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "asap"
OTHER INFORMATION: /strand.name=
OTHER INFORMATION: "7.7.7.7.7.8.9.8.9.5"
US-08-474-633A-74

Query Match 3.8%; Score 53.4; DB 1; Length 243;
Best Local Similarity 53.0%; Pred. No. 9.3e-07;
Matches 114; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 634 ATGAGCCCAACAGGGGATATGAGCCCACTATGGAATGAGCCCGCTGTGG 693
DB 2 ATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAG 61
QY 694 GATATGAGTCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 753
DB 62 GCGATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTG 121
QY 754 GGGATATGAGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACT 813
DB 122 AAGCGATGAGAGAGAGCTTAAGAGAGATGAGAGAGAGCTGAAGATGAGAGAGAG 181
QY 814 CTGGGGATATGAGAGCCCACTGAGAGATATGAG 848
DB 182 CTCAGAGATGAGAGAGAGCTTAATGATGAG 216

RESULT 15

US-08-823-771-74
Sequence 74, Application US/08823771
Patent No. 6459019

GENERAL INFORMATION:

APPLICANT: E. I. DU PONT DE NEMOURS AND
COMPANY

TITLE OF INVENTION: CHIMERIC GENES AND
METHODS FOR INCREASING
INCREASING THE LYSINE
AND THREONINE CONTENT

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:

ADDRESSER: E. I. DU PONT DE NEMOURS
AND COMPANY

STREET: 1007 MARKET STREET

CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: U.S.A.

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MICROSOFT WORD VERSION 2.0C

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/823,771

FILING DATE: 24-Mar-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/474,633

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: BARBARA C. STEGELL

REGISTRATION NUMBER: 30,684

REFERENCE/DOCKET NUMBER: BB-1037-C

TELEPHONE: 302-992-4931

TELEFAX: 302-773-0164

TELEX: 835420

INFORMATION FOR SEQ ID NO: 74:

SEQUENCE CHARACTERISTICS:

LENGTH: 243 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: E. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: 2-9

FEATURE:

NAME/KEY: CDS

LOCATION: 2..235

OTHER INFORMATION: /function= "synthetic"

storage protein

/product= "protein"

/gene= "asap"

/strand.name=

"7.7.7.7.7.8.9.8.9.5"

SEQUENCE DESCRIPTION: SEQ ID NO: 74:

US-08-823-771-74

Query Match 3.8%; Score 53.4; DB 4; Length 243;
Best Local Similarity 53.0%; Pred. No. 9.3e-07;
Matches 114; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 634 ATGAGCCCAACAGGGGATATGAGCCCACTATGGAATGAGCCCGCTGTGG 693
DB 2 ATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAG 61
QY 694 GATATGAGTCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 753
DB 62 GCGATGAGAGAGAGAGCTGAAGCGATGAGAGAGAGCTGAAGCGATGAGAGAGAGCTG 121
QY 754 GGGATATGAGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACT 813
DB 122 AAGCGATGAGAGAGAGCTTAAGAGAGATGAGAGAGAGCTGAAGATGAGAGAGAG 181
QY 814 CTGGGGATATGAGAGCCCACTGAGAGATATGAG 848
DB 182 CTCAGAGATGAGAGAGAGCTTAATGATGAG 216

Search completed: December 16, 2003, 17:36:27
Job time: 103.547 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 16, 2003, 10:43:10 Search time 465.02 Seconds

(without alignments)
10099.015 Million cell updates/sec

Title: US-09-864-291-4

Perfect score: 1413
1 cgccagcagggcgccagca.....gcataaaaaaaaaaaaaa 1413

Sequence: IDENTITY NUC

Scoring table: Gapop 10.0, Gapext 1.0

Searched: 2201672 seqs, 166179599 residues 4403344

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	262.6	18.6	467	11	US-09-764-891-1230
2	241.2	17.1	436	9	US-09-864-761-31561
3	220	15.6	7099	11	US-09-764-891-7570
4	195.4	13.8	321	11	US-09-764-891-1356
5	122	8.6	432	11	US-09-918-995-13536
6	116.4	8.2	409	10	US-09-960-352-6984
7	92.4	6.5	291	11	US-09-764-891-7571
8	86	6.1	471	9	US-09-864-761-15029
9	78.6	5.6	511	13	US-10-029-386-24981
10	77.8	5.5	593	13	US-10-027-632-290832
11	77.8	5.3	593	14	US-10-027-632-290832
12	75.2	5.3	420	10	US-09-974-300-3626
13	69.4	4.9	866	13	US-10-029-386-22839
14	68.8	4.9	591	9	US-09-864-761-19727
15	68.8	4.9	1959	9	US-09-864-761-2946

16	68.6	4.9	2016	10	US-09-938-842A-2004	Sequence 2004, Ap
17	64.4	4.6	5769	11	US-09-764-891-7572	Sequence 7572, Ap
18	63.6	4.5	1040	13	US-10-029-386-26033	Sequence 26033, A
19	61.6	4.4	541	13	US-10-029-386-11265	Sequence 11265, A
20	59.8	4.2	3489	13	US-10-294-804-1	Sequence 1, Appl1
21	59.4	4.2	541	13	US-10-029-386-12333	Sequence 12333, A
22	58.8	4.2	417	9	US-09-864-761-27093	Sequence 27093, A
23	58.8	4.2	3048	13	US-10-027-632-113439	Sequence 113439, A
24	58.8	4.2	3048	14	US-10-027-632-113439	Sequence 113439, A
25	57.2	4.0	511	13	US-10-029-386-24981	Sequence 24981, A
26	55.6	3.9	628	13	US-10-029-386-22859	Sequence 22859, A
27	54.8	3.9	2108	10	US-09-962-832-225	Sequence 225, App
28	54.6	3.9	417	9	US-09-864-761-27093	Sequence 27093, A
29	54.6	3.9	422	10	US-09-738-973-337	Sequence 337, App
30	54.6	3.9	422	10	US-09-854-133-337	Sequence 337, App
31	54.6	3.9	422	15	US-10-144-649A-337	Sequence 337, App
32	54.6	3.9	3048	13	US-10-027-632-113439	Sequence 113439, A
33	54.6	3.9	3048	14	US-10-027-632-113439	Sequence 113439, A
34	53.8	3.8	541	13	US-10-029-386-11265	Sequence 11265, A
35	53.4	3.8	243	15	US-10-023-066A-74	Sequence 74, Appl1
36	53.2	3.8	3773	9	US-09-925-302-47	Sequence 47, Appl1
37	53.2	3.8	6354	15	US-10-084-817-158	Sequence 158, App
38	52.6	3.7	354	9	US-09-864-761-18194	Sequence 18194, A
39	52.6	3.7	421	13	US-10-029-386-14222	Sequence 14222, A
40	52.6	3.7	597	13	US-10-029-386-517	Sequence 517, App
41	52.6	3.7	2030	10	US-09-880-107-2416	Sequence 2416, App
42	52.6	3.7	8923	10	US-09-764-847-1041	Sequence 1041, App
43	52.6	3.7	8923	15	US-10-092-154-1041	Sequence 1041, App
44	52.4	3.7	2746	13	US-10-371-725-1	Sequence 1, Appl1
45	52.2	3.7	689	13	US-10-027-632-148124	Sequence 148124, A

ALIGNMENTS

RESULT 1
US-09-764-891-1230
Sequence 1230, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1230
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: SITE
LOCATION: (421)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1230

Query Match	18.6%, Score 262.6, DB 11, Length 467, Best Local Similarity 76.4%, Pred. No. 4.6e-74, Matches 365, Conservative	0, Mismatches 100, Indels 13, Gaps 3,
DB	386 CAGGAAGAAGAGTCCATCGAATTCGCACTGATGTAAAGCTGCTGCTGCTC 445	
QY	2 CAGGAAGAAGAGTCCATCGAATTCGCACTGATGTAAAGCTGCTGCTGCTC 61	
DB	446 CAGGAATTCACCTGGAAGTGAATTCGCTGGAACCTCAGACCTGATCATAT 505	
QY	62 CCGAAGATTCACCTGGAAGTGAATTCGCTGGAACCTCAGACCTGATCATAT 121	
DB	506 TACTGTCACAGAGGCTCAGTGTCTCTCAACAGACCTGTCACATATTCAT 565	
QY	122 TACTGTCACAGAGGAT---ATGTGACATCCACAGATGCTTGTTCAG-----TTAT 169	

```

Qy 566 GATCTATGAGCCCCCAGCAGATATACAGTCCAAACGAGGGAATATGAACTCCACC 625
Db 170 TGTCTATGAGCCCCCAGCAGATATGAGAGCCCACTCCCGGATACGAGAGCCCAACC 229
Qy 626 AGAAGATATGAGCCCAACGAGGGAATATGAGAGCCCACTCTATGAGATATGAGAGCC 685
Db 230 TGCAGATATGAGAGCCCAACGAGGGAATATGAGAGCCCACTCTATGAGATATGAGAGCC 289
Qy 686 GCGTGTGAGATATGAGAGTCCCACTGAGGGAATATGAGATCCCACTGAGGGAATATGAGT 745
Db 290 ACTGTGCAATATGAGAGCCCACTTGTGATACGAGAGCCCACTGAGGGAATATGAGAGC 349
Qy 746 CCCACCTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGATCCCACTGAGGGAATATG 804
Db 350 CCCACCTGAGGGAATATGAGAGCCCACTTGTGATACGAGAGCCCACTGAGGGAATATG 409
Qy 805 GAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGATCCCACTGAGGGAATATG 862
Db 410 GAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATG 467

```

RESULT 2

```

US-09-864-761-31561
/ Sequence 31561, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aeonica-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ CURRENT FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117

```

```

/ SOFTWARE: Annonex Sequence Listing Engine vers. 1.1
/ SEQ ID NO 31561
/ LENGTH: 436
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ PRIMER:
/ OTHER INFORMATION: MAP TO 299716.4
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
/ OTHER INFORMATION: EST HUMAN HIT: AM169980.1, EVALUE 8.00e-11
/ OTHER INFORMATION: SWISSPROT HIT: P18616, EVALUE 3.00e-13
US-09-864-761-31561

```

```

Query Match 17.1%; Score 241.2; DB 9; Length 436;
Best Local Similarity 73.2%; Pred. No. 3.8e-67;
Matches 328; Conservative 0; Mismatches 108; Indels 12; Gaps 1;

```

```

Qy 562 TTGTGATCTATGAGAGCCCAACGAGGGAATATGAGAGCCCACTCTATGAGATATGAG 621
Db 1 TTATGTCTATGAGAGCCCACTGAGGGAATATGAGAGCCCACTCTCTGAGATACGAG 60
Qy 622 CACGAGAGATATGAGAGCCCAACGAGGGAATATGAGAGCCCACTCTATGAGATATGAG 681
Db 61 CACTGAGAGATATGAGAGCCCAACGAGGGAATATGAGAGCCCACTCTGAGATACGAG 120
Qy 682 CCCCCCTGTGAGATATGAGATCCCACTGAGGGAATATGAGATCCCACTGAGGGAATATG 741
Db 121 CCTCAGCTGTGAGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTCTGAGATATG 180
Qy 742 GAGTCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGATCCCACTGAGGGA 801
Db 181 GAGCCCACTCTGAGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTCTGAGAT 240
Qy 802 ATGTGTCCCACTGAGGGAATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAG 861
Db 241 ATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAG 300
Qy 862 GAATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAG 921
Db 301 GATACAGAGCCCACTCTGAGATATGAGAGCCCACTGAGGGAATATGAGAGCCCACTGAG 356
Qy 922 ACAGATCTATGAGAGCCCACTGAGGGAATATGAGAGCCCACTCTGAGATATGAGAG 981
Db 357 -----CAGGCTCTGAGGGAATATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAG 408
Qy 982 ATTACCACTCTGAGGGAATATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAG 1009
Db 409 ATTCTTAACCTCTGAGGGAATATGAGAGCCCACTCTGAGATATGAGAGCCCACTGAG 436

```

RESULT 3

```

US-09-764-891-7570
/ Sequence 7570, Application US/09764891
/ Publication No. US20030077808A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PC006
/ CURRENT APPLICATION NUMBER: US/09/764,891
/ PRIOR FILING DATE: 2001-01-17
/ PRIOR APPLICATION data removed - consult PALM or file wrapper
/ NUMBER OF SEQ ID NOS: 10231
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 7570
/ LENGTH: 7099
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-764-891-7570

```

```

Query Match 15.6%; Score 220; DB 11; Length 7099;
Best Local Similarity 75.3%; Pred. No. 1.4e-59;

```

Matches 274; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 554 ATATCCATTGTATGATATGAGACCCACACAGATATACAGTTCACACAGGAGATA 613
Db 6718 ATTCCTGATATGTATGTATGAGACCCACCTGACAGATATGAGACCCACCTCCGAGATA 6777
QY 614 TGAAGCTCCACAGAGATATGAGACCCACAGGAGATATGAGACCCACCTATAGG 673
Db 6778 CGAGACCCACCTGACAGATATGAGACCCACCTGATAGAAATAGAGCCCGCTGAGG 6837
QY 674 ATATGAGACCCCGCTGTGAGATATGAGTCCACCTGAGGAGATATGAGTCCACCTG 733
Db 6838 ATACAGAGCTTCACCTGTGAGATATGAGCCCACTCTGAGATACAGAGCCCACTG 6897
QY 734 GAGATATGAGTCCACCTGTGAGATATGAGCCCACTGTGAGGATATGAGTCCAC 793
Db 6898 AGATATGAGACCCACCTCTAGATATGAGACCCCACTTGTATATGAGACCCAC 6957
QY 794 TGGGAGATATGAGTCCCACTGTGAGATATGAGACCCCACTGACAGATATGAGCC 853
Db 6958 TCTGAGATATGAGACCCCACTCTGAGATATGAGACCCCACTGACAGAAATAGAGCC 7017
QY 854 ACCAGCTGAAATGAGCCCTACCTGACATATGAGCTTCATGCTGAGAAATACAG 913
Db 7018 GCTGCGGAGATACAGAGCTTCACCTGTGATCAGAGCCAGCCCTCAGAGATCTACAG 7077
QY 914 TGCC 917
Db 7078 AGCC 7081

RESULT 4

US-09-764-891-1356
; Sequence 1356, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIORITY FILING DATE: 2001-01-17
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1356
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: SITE
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1356

Query Match

Best Local Similarity 13.8%; Score 195.4; DB 11; Length 321;
Matches 246; Conservative 2; Mismatches 69; Indels 1; Gaps 1;

QY 64 GCCCTGCTGGAGCCCTCATCCCTCTGCGGAAAGTGTCTGAGAGCTGAGAGATG 123
Db 1 GCCCGCCGAGGAGCCCTCATCCCTCAACGATGAAGTCTCTGAGAGCGGTCTCCGAGATG 60
QY 124 ACCCTGCTTCTTCAAGAAACAGTGAATCTTCTCTTAATGAGCAAGAAAGAA 183
Db 61 AGCTCTCTTCCACAGAGATCAGAGGCTCAATGCTTTTATGAGTGAAGACAGAA 120
QY 184 CGTTGTTCTCACTTACAGGAGTGTCTTGTGACTTCACTTATGTCATAGTACCCCA 243
Db 121 CATGTTTCTCACTTACAGGAGTGTCTTATGCTTCACTGCTCATGATGCCA 180

QY 244 TGTCTTCTTATGATGCGCTTGTGCTGATGATGATGATGATGATGATGATGATGAT 303
Db 181 TGTGCTTCTTATGATGCGCTTGTGATGATGATGATGATGATGATGATGATGAT 240
QY 304 TTGCCCCCACTATTAAGAAACATTCAGGAGCTTCAAGGTGTGCTGAG-ANGA 362
Db 241 TTCTGCAACCTTCAATTAAGGAACTATTCAGGAGCTTCAATGATGAGGCTGAGAAAG 300
QY 363 CAAGCTGTTTAAAGTA 380
Db 301 CAAGCTATCTTAAAGTA 318

RESULT 5

US-09-918-995-33536
; Sequence 33536, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIORITY FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIORITY FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33536
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc_feature
; LOCATION: (1)...(432)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

Query Match

Best Local Similarity 8.6%; Score 122; DB 11; Length 432;
Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;

QY 36 ATGCAAGTGAACAGAGCCACACGAGAGCGGTGAGGAGCCCTACCCCTGAGGAA 95
Db 18 ATGCGCTCAACAGATATCTGAGAGCGGCGAGATATCTCAATTAAC-CAAG 74
QY 96 AGTCTTGAAGAGTGAAGATGAGATGAGCTCTGCTCTTACAGAAACAGTGAATCC 155
Db 75 AGCATCTATAGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 134
QY 156 TATCTCTTATATGCAACAAAGAAAGAGATGTTTCTCACTTCAATACGAGGTCTTC 215
Db 135 GAAGCTTCAAGAGAGCAAGAAAGCACTGTCTTACCTTACCGGCTCATCTTT 194
QY 216 GAGCTTCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 275
Db 195 CTGTC-----CAAGGCAAGATGATGATGATGATGATGATGATGATGATGATG 248
QY 276 AGTATGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 335
Db 249 AAGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 308
QY 309 GCGAAGCGGAGAGTGTGAGAGGCTGTGCTTCAAGATGATGATGATGATGATG 368
Db 396 GGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 451
QY 369 GCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 424

RESULT 6

US-09-960-352-6984

Sequence 6984, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Mairen, Wesley C.
APPLICANT: Tao, Mengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE OF INVENTION: MOSCIE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 6984
LENGTH: 409
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 30-LIB3058-024-Q1-K1-H9
US-09-960-352-6984

Query Match 8.2%; Score 116.4; DB 10; Length 409;
Best Local Similarity 59.2%; Pred. No. 9.6e-27;
Matches 239; Conservative 0; Mismatches 156; Indels 9; Gaps 2;

QY 26 CCTGGGCGAGATGAGTGAACCAAGCCAGACCCGCTGCGGGCCCTCATGCC 85
DB 13 CCGAGGAGCTATGGCTCTCATGAAGCACTCGAGGGCGGCGAGTATGCTCAACAA 72
QY 86 CTCTGGCGAAGGTCTTGAAGCAGTGAAGTGAAGTGAAGTCTTCTGCTCTTACAGAAAC 145
DB 73 CAC--CGAGAGATCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCT 129
QY 146 AGTGAATCTTATCTTATATGCAAAAGAAAGAAAGTGTCTTCTCATTAACCG 205
DB 130 GAATGCGCAAGAGCCCTTCAAGGGGCAAGAAAGGACCGTCTTACCTTACCCGTAACG 189
QY 206 GGTGGTCTTCTGACCTTCACTTATGCTTATGCTTATGCTTATGCTTATGCTTATGCT 265
DB 190 GGTGATCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 243
QY 266 TGGCGATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 325
DB 244 CTATCTGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 303
QY 326 AACCATTCAGGAGCTTCAAGTGTGCTGAGGAGGAGCAAGCTGTTTAAAGTATCTT 385
DB 304 GACAGTGAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 363
QY 386 CAGGAAAGAGTGCATGCAATTTGCCCCAATGATGTAAG 429
DB 364 TATGTTGCGGGGCGCATGCAATATGACAGCGGATGTTACAG 407

RESULT 7
US-09-764-891-7571
Sequence 7571, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7571
LENGTH: 291
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-7571

Query Match 6.5%; Score 92.4; DB 11; Length 291;

Best Local Similarity 76.0%; Pred. No. 4.6e-19;
Matches 114; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 206 GGTGGTCTTCTGACCTTCACTTATGATACCCCAATGCTTCTTATGATGCCGT 265
DB 9 GGTGATTTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT 68
QY 266 TGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 325
DB 69 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 128
QY 326 AACCATTCAGGAGCTTCAAGTGTGCTG 355
DB 129 AACATTCAGGAGCTTCAATGATGATG 158

RESULT 8
US-09-864-761-15029
Sequence 15029, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Mengheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
SEQ ID NO 15029
LENGTH: 471
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

OTHER INFORMATION: MAP TO Z99716.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
US-09-864-761-15029

Query Match 6.1%; Score 86; DB 9; Length 471;
Best Local Similarity 77.6%; Pred. No. 7.3e-17;

Matches 104; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 662 CCCACCTATGAGATGAGAGCCCGCTGTGGATATGAGTCCACCTGGGGATATG 721
DB 338 CCCACTATGCTATGAGAGCCCGCTGTGGATATGAGTCCACCTGGGGATATG 397
QY 722 AGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCCACCTGGGGATATG 781
DB 398 AGCCCGCTGTGGATATGAGAGCCCGCTGTGGATATGAGTCCACCTGGGGATATG 457
QY 782 TGGAGTCCACCTG 795
DB 458 CAGAGCTCACCTG 471

RESULT 9

US-10-029-386-24981
Sequence 24981, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
FILE REFERENCE: ABOVICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annonex Sequence Listing Engine vers. 1.1
SEQ ID NO 24981
LENGTH: 511
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO Z97205.1
US-10-029-386-24981

Query Match 5.6%; Score 78.6; DB 13; Length 511;
Best Local Similarity 49.9%; Pred. No. 1.9e-14;

Matches 198; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 520 CTGACGTGCTCTTCAACAGACCTTGTCCAGCATATTCATATGATATGAGCC 579
DB 106 CTGACGAGCCCTCTCTTCAACAGAGTCCCTCTCTCTCACTGAGGCTCCCTCTC 165
QY 580 CACCAACAGATATACAGTCAACCAAGGGAATATGAACTCCACCAAGAGATATGAG 639
DB 166 CACCTGAGGCTCCCTCTCTCAACCTGAGGCTCCCTCTCTCACTGAGGCTCCCTC 225
QY 640 CCAACAGAGGATATGAGAGCCCACTTATGAGATATGAGAGCCCGCTGTGGATATG 699
DB 226 CTGACCTGAGAGTCTCTCTCTCAACCTGAGGCTCCCTCTCTCACTGAGGCTCT 285
QY 700 GAGTCCACCTGGGGATATGAGTCCACCTGGGGATATGAGTCCACCTGGGGAT 759
DB 286 CTGCTTCACTGAGGCTCTCTCTCTCACTGAGGCTCTCTCTCTCACTGAGGCT 345
QY 760 ATGAGAGCCCACTGGGGATATGAGTCCACCTGGGGATATGAGTCCCACTGGGG 819
DB 346 CCTCTCTCACTGAGGCTCTCTCTCTCAACCTGAGGCTCTCTCTCTCACTGAGG 405
QY 820 GATATGAGAGCCCACTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTGAG 879

DB 406 GCTCCCTCTCTCACTGAGGCTCTCTCTCTCACTGAGGCTCCCTCTCTCACTG 465
QY 880 CTGATATGAGAGCTCTCTCTCTGGAATATGAGTCC 916
DB 466 CAGGCTCCCTCTCTCACTGAGGCTCTCTCTCTCC 502

RESULT 10

US-10-027-632-290832
Sequence 290832, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 290832
LENGTH: 593
TYPE: DNA
ORGANISM: Human
US-10-027-632-290832

Query Match 5.5%; Score 77.8; DB 13; Length 593;
Best Local Similarity 54.9%; Pred. No. 3.8e-14;

Matches 151; Conservative 1; Mismatches 123; Indels 0; Gaps 0;

QY 177 AAAGGAGCTGTTTCTCATCTTACCGGGTGTCTGTGACTTCACTTATGAT 236
DB 45 AAAGGAGCAAGAAAGGACATCTTACCTTACCTTCTTCTGTCAGGGCGAAG 104
QY 237 GACCCATGCTTTCTTTATGATGCGCTTGTGAGTATGATGATGATGATGATG 296
DB 105 GATGCAAGGAGTCTTGTATGATGATGATGATGATGATGATGATGATGATG 164
QY 297 CCAATTTTGGCCCCCACTATATTAAGAAACATTCAGGAGCTCCAGGTGGTGG 356
DB 165 CTTGATTTGACAACTGATCAAGGAAACATGAAACACCAAGAGGCGGTGG 224
QY 357 GAGGACAAGCTGTTTATGATATCTTCAAGAAAGAGGTGTCATGAAATTTG 416
DB 225 GAGGCTGCTGTTTGGGAGGCTTCAAGGAGGCGGCGGCACTGAAATTTGAG 284
QY 417 CTGATGTAAAGCTGCTGCTGCTGCTGCGAGG 451
DB 285 CAGGTCTCCAGGTGATCTCAAGGCTTCAAGG 319

RESULT 11

US-10-027-632-290832
Sequence 290832, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome

FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 290832
LENGTH: 593
TYPE: DNA
ORGANISM: Human
US-10-027-632-290832

Query Match 5.5%; Score 77.8; DB 14; Length 593;
Best Local Similarity 54.9%; Pred. No. 3.8e-14;
Matches 151; Conservative 1; Mismatches 123; Indels 0; Gaps 0;

QY 177 AAGGAACTGTTTCTCACTTCAATCCGAGTGTCTTCTGACTTCACTTCAAT 236
DB 45 AAGGAGCCAAAGGACACATCTTACCCCTTCACTTCTTCTGTCAGAGGAG 104
QY 237 GACCCATGCTTCTTTATGATGCGTTGGCCTGATGATGATGATGATGATGAT 296
DB 105 GATGCGACGAGTCTCTGTAATGATCTTTATCTTGAAGGCTTATGATCAAGCAG 164
QY 297 CCAATTTTCCCACTACATTAAGAACATTCAGGACGCTCAGGTGCTGCTG 356
DB 165 CCGTGTGTTACCAACATCAGTCAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 224
QY 357 GAAGGACAGCTGTTTAAATGATCTTCAAGGAAAGGAGTGCATGCAATTTGCCCA 416
DB 225 GAAGGCTCTCTTCYGGCAGTCCGCTTCAAGGAGGAGGAGGAGGAGGAGGAG 284
QY 417 CTGATGCTAAAGTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 451
DB 285 CAGTGTCTCCAGGATGATCTCAAGGCTGCAAGGCTGCAAGGCTGCAAGGCT 319

RESULT 12
US-09-974-300-3626
Sequence 3626, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groch
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3626
LENGTH: 420
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-3626

Query Match 5.3%; Score 75.2; DB 10; Length 420;
Best Local Similarity 49.7%; Pred. No. 2.1e-13;
Matches 191; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

QY 557 TCCAAATGATCTATGACCCCAACACAGATATCAATCAACAGAGGAAATATG 616
DB 8 TCGAAGCGGCTATATGAGTCAAGGAGGCTGATATGAGTCAAGGAGGCTGATATG 67
QY 617 AACTCCACCAAGAGATATGAGCCCAACAGGAGGATATGAGCCCACTTATGATAT 676
DB 68 CAGTCAAGCGGCTATATGAGTCAAGGAGGCTGATATGAGTCAAGGAGGCTGATAT 127
QY 677 TGAAGCCCGCTGTGAGATATGAGTCCCACTTGGGAGATATGAGTCCCACTTGG 736
DB 128 TGAAGTCAAGCGGCTGATATGAGTCAAGGAGGCTGATATGAGTCAAGGAGGCTG 187
QY 737 ATATGAGTCCCACTTGGGAGATATGAGCCCACTTGGGAGATATGAGTCCCACTTGG 796
DB 188 ATATGAGTCCCACTTGGGAGATATGAGCCCACTTGGGAGATATGAGTCCCACTTGG 247
QY 797 GAGATATGAGTCCCACTTGGGAGATATGAGCCCACTTGGGAGATATGAGTCCCACT 856
DB 248 CTGATATGAGTCCCACTTGGGAGATATGAGCCCACTTGGGAGATATGAGTCCCACT 307
QY 857 AGTGAATATGAGCCCTTACCCCTGCAATATGAGTCCATCTGCTGGAATATGAGTCC 916
DB 308 CCGCTGATATGAGCCCTGCAATATGAGTCCATCTGCTGGAATATGAGTCCATCTG 367
QY 917 CTCTCAAGATCATATGAGCTCA 940
DB 368 AACCGATGATATGAGCTGCTAA 391

RESULT 13
US-10-029-386-22839/C
Sequence 22839, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AEOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 22839
LENGTH: 866
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
OTHER INFORMATION: MAP TO AC005630.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
OTHER INFORMATION: SWISSPROT HIT: 008379, EVALUOR 4.00e-95
OTHER INFORMATION: NT HIT: g116161055, EVALUOR 4.00e-95
OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALUOR 0.00e+00
US-10-029-386-22839

Query Match 4.9%; Score 69.4; DB 13; Length 866;
Best Local Similarity 49.9%; Pred. No. 2.5e-11;
Matches 175; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 571 ATGACCCCAACCAAGATATATGATCAATCAAGGAGGAAATATGAACTCCACCAAG 630
DB 509 ACCTAGCTCTCTCTCTGTTCAACAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 450
QY 631 GATATGAGCCCAACCAAGGAGATATGAGCCCACTTATGATATGAGCCCACTT 690

Job time : 469.02 secs

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2946
; LENGTH: 1959
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002041.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 4.3
US-09-864-761-2946
```

```

Query Match 4.9%; Score 68.8; DB 9; Length 1959;
Best Local Similarity 51.6%; Pred.No.6.2e-11;
Matches 157; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
```

```

QY 602 ACCAGGGGATATGGAATCCAGCAGAGATATGAGCCCAACGAGGGGATATGAGC 661
   |||||
DB 479 ACCTGGGGTATGTGAGGTGACGTGGGAATGTGTGAGTGACCTGGGGGACGTGTAG 538
   |||||
QY 662 CCACCTATGGGATATGAGCCCGCTGTGGGATATGAGTCCACCTGGGGGATATGG 721
   |||||
DB 539 GTGACCTGTGGGACATGTGAGGTGACCTGGGGATGTGTGAGGTGACCTGGGGGACGTGT 598
   |||||
QY 722 AGTCCACCTGGGGGATATGAGTCCACCTGTGGGGATATGAGGCCCACTGGGGGATA 781
   |||||
DB 599 GAGGTGACCTGGGGATGTGTGAGGTGACCTGTGGGACATGTGAGGTGACGTGTGACATG 658
   |||||
QY 782 TGGAGTCCACCTGGGGGATATGTGCCCCACCTGGGGGATATGAGGCCCACTGTGCAGG 841
   |||||
DB 659 TGTGAGGTGACGTGGGGATGTGTGAGGTGACCTGGGGGACGTGTGACGTGACGTGGGGA 718
   |||||
QY 842 ATATGAGCCCAACCACTGGAATGAAGCCCTACCCCTGCATATGAAGCTCCATCTGC 901
   |||||
DB 719 ATGTGTGAGGTGACCTGGGGGACGTGTGAGGTGACCTGGGGGATGTGTGAGGTGACCTGG 778
   |||||
QY 902 TGGG 905
   |||
DB 779 GGGG 782
```

Search completed: December 16, 2003, 17:58:42

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:21:45 ; Search time 2907.96 Seconds

(Without alignments)
11809.735 Million cell updates/sec

Title: US-09-864-291-4

Perfect score: 1413

Sequence: 1 cggcagcgagggcgcgagga.....gcataaaaaaaaaaaaaa 1413

Scoring table: IDENTITY_NTC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba:*
2: em_estbm:*
3: em_estbn:*
4: em_estbu:*
5: em_estbv:*
6: em_estbp:*
7: em_estbr:*
8: em_estbs:*
9: gb_estl:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: em_estfm:*
16: em_estfn:*
17: em_estfu:*
18: em_estfv:*
19: em_estfw:*
20: em_estfx:*
21: em_estfy:*
22: em_estfz:*
23: em_estga:*
24: em_estgb:*
25: em_estgc:*
26: em_estgd:*
27: em_estge:*
28: gb_estl:*
29: gb_est2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	546.8	38.7	1079	12	BM564167 AGENCOURT
2	422.4	29.9	1414	11	AK015863 Mub mubcu
3	346.2	24.5	663	10	BC699398
4	345.4	24.4	775	10	BG722815

5	341.8	24.2	717	14	BY715414
6	211.4	15.0	701	10	BG701881
7	197.8	14.0	360	13	BY098622
8	183.6	13.0	932	13	BU107273
9	182.2	12.9	814	13	BU961805
10	179.4	12.7	500	13	BU103740
11	179.4	12.7	760	13	BU284977
12	179.4	12.7	723	9	AJ452591
13	177.8	12.6	568	12	BU390687
14	177.8	12.6	636	12	BM490729
15	167.4	11.8	772	9	AJ454950
16	166.8	11.8	706	13	BU433499
17	155.4	11.0	460	10	BB680329
18	155	11.0	719	14	CB593075
19	154	10.9	760	14	CB592279
20	153.8	10.9	862	13	BU735296
21	153.8	10.9	868	14	CD255258
22	153.6	10.9	931	13	BU916890
23	153.6	10.9	931	13	BU914288
24	152.2	10.8	613	9	AL849211
25	152.2	10.8	647	9	AL849651
26	152.2	10.8	655	9	AL898723
27	152.2	10.8	661	9	AL972453
28	152.2	10.8	663	9	AL881810
29	152.2	10.8	663	9	AL896682
30	152.2	10.8	726	14	CB591559
31	152	10.8	509	10	BS15416
32	152	10.8	594	12	BM190972
33	152	10.8	594	12	BM190975
34	151.6	10.7	460	10	BS13187
35	150.6	10.7	611	9	AL846063
36	150.4	10.6	882	14	CD301781
37	143.6	10.2	488	28	AQ997933
38	141.2	10.0	635	9	AL889954
39	136.8	9.7	622	10	BE468866
40	135.6	9.6	899	10	BE470466
41	132.8	9.4	652	14	CA590725
42	131.8	9.3	631	13	BU805447
43	131.8	9.3	744	13	BU230211
44	131.2	9.3	635	13	BU807307
45	130	9.2	437	13	BY249862

ALIGNMENTS

RESULT 1
LOCUS BM564167 1079 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT 6560149 NIH_MGC_119 Homo sapiens CDNA clone IMAGE:5742129
5' mRNA sequence.

ACCESSION BM564167
VERSION BM564167.1 GI:18811738
KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE
1 (bases 1 to 1079)
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

JOURNAL
TITLE NIH-MGC http://mgs.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12759 row: m column: 10
High quality sequence stop: 726.

FEATURES
SOURCE

Location/Qualifiers
1. 1079
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5742129"
/cissue_type="medulla"
/lab_host="DH10B"
/clone_11b="NIH-MGC_119"
/note="Organ: brain; Vector: PCMV-SPORT6; Site: 1; NotI; Site: 2; EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb. Insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (invitrogen). Research Genetics tracking code 013. Note: this is a NIH-MGC Library."

BASE COUNT 267 a 297 c 268 g 247 t
ORIGIN

Query Match 38.7%; Score 546.8; DB 12; Length 1079;
Best Local Similarity 73.7%; Pred. No. 4.2e-113;
Matches 772; Conservative 0; Mismatches 252; Indels 24; Gaps 5;

21 GGGGCGCTGGGCGAGTGAACGAGCCACACCGAGAGCGCTGCGGGCCCTC 80
22 GAGGCGCCGAGAGAGAGCGGTGATGAGCCACCGAGAACCGCGCGAGCCCTC 91
81 ATCCCTCTGCGGAGAGTCTTTGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 140
92 ATCCCTTAACGAGAGAGTCTTTGAGAGCGGTCTCGAGTGTGAGAGTGTGAGAG 151
141 AAACGAGTGAATCTATCTCTTATGAGCAAGAGAGAGAGAGAGAGAGAGAGAG 200
152 CGATGAGAGAGTCTTAATCTCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAG 211
201 TAACGCGGTGCTTCTGAGCTTCACTTATGAGAGAGAGAGAGAGAGAGAGAGAG 260
212 TACCGGAGTATTTTCACTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 271
251 CCGTGTGAGTGTGAG 320
272 CCATTGATGATGAG 331
321 AAAGAGCAATCAG 380
332 AAGGAGATGATGAG 391
381 TCCCTTCAAG 440
392 GTCTTCAAG 451
441 GTCGCAAG 500
452 GCTGCGCAAG 511
501 ATAAATTAATGAG 560
512 GTAATTAATGAG 560
561 ATTGATGATGAG 620
561 -TTATGATGATGAG 619
621 CCACGAG 680
620 CCACGAG 679
681 GCGCGCGCTGAG 740
680 GCTTCACTGAT 739

QY 741 GAG-TGCCACTGAGAGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 799
DB 740 GAG 799
QY 800 ATATGAG 859
DB 800 ATATGAG 859
QY 860 TGAAG 919
DB 860 GAGATGAG 912
QY 920 TCAAGATGATGAG 979
DB 913 ---GAGGCGAG 969
QY 980 CCATTTCACAG 1038
DB 970 CCATTTCACAG 1029

QY 1039 AACTGAGTCAAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1066
DB 1030 AATTGAG 1057

RESULT 2
AK015863
LOCUS
DEFINITION
AK015863 1414 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930521I23 product:hypothetical protein, full insert sequence.
AK015863
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus (house mouse)

REFERENCE
AUTHORS
TITLE
JOURNALS
MEDLINE
PUBMED
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
10349636
3
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
11042159

REFERENCE
AUTHORS
TITLE
JOURNALS
MEDLINE
PUBMED
11076861
4
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S., Aizawa, K., Iizawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamana, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., Kling, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,

Oy	780	TATGGAAGTCCCACTCGAGGGGAATATATGTGCCCCACCGGGGAATATAGAGCCCCACATGCA	839
Db	775	TATGGAAGTCCCACTCTCGAATATGACCTTCCACTGTATGATAGGAATCCCCACTCTCT	834
Oy	840	GAATATGAGCCCCCACTGACCTGGAATGAGGCTTACCCCTCGCATATGAGCTCATCT	899
Db	835	GGATATGAGGCCCACTATGAGATATGAGGCCCACTCCAGATACGAACTACACTT	894
Oy	900	GCTGGAAATACAGCTGCTCTCACAGATCT	929
Db	895	ATGGATCCGATTCCCCCACTCCAGATAT	924

RESULT 3	663 bp	mRNA	linear	EST 07-MAY-2001
LOCUS	602679081P1	NTH_MGC_95	Homo sapiens	CDNA clone IMAGE:4811804 5',
DEFINITION	mRNA sequence.			

VERSION	BG699398.1	GI:13967653
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Bumaiyot, Metazoa; Chordata; Crustacea; Vertebrata; Rutelesocoma; Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
1 (bases 1 to 653)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov
Tissue Procurement: M.I.K.O.S. Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroaki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLNL0703 row: a column: 21
High quality sequence: sncp: 660.

FEATURES	Location/Qualifiers
SOURCE	1. .663

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:481804"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_id="N1H_MGC_95"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-ttttttttttttttn-3',
size-selected for average insert size 2.5 kb and
normalized to RQF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NIHRI, National
Institutes of Health). Note: this is a N1H_MGC library."

```

Query Match	24.5%	Score 346.2	DB 10	Length 663
Best Local Similarity	74.7%	Pred. No. 9.8e-68		
Matches 490; Conservative	0	Mismatches 151	Indels 15	Gaps 4

QY 21 GGGGGCTGGGCAAGATGGCAGTGCAGTGAACCGAGCCACCGAGAGGCGTGTGATGGGGCCCTC 80
Db 22 GAGGCCCGAAGCAAGATGGCGGTGATCAGAGCCACCGAGAACCGCGCGGAGGCCCTC 81
QY 81 ATCCCTCTGGGAAAGTGTCTTGAAGCAGGTGAGATGATGGACCTGTGCTTCTACAG 140

Db	82	ATCCCTAA	CGGTGAAGAAGCTCTTGAAAGGGGCTCCGAATGAGAGCTCTCTCCACAG	141
Qy	141	AAACGAGTGAAT	CTCTATCTCTTTAATGGCACAAAGAAAGAAAGCTGTTCTCACTTCA	200
Db	142	CGATCAGAAAGGCTCAAA	TGCTCTTAAAGTGAAGAAAGCAGAGAC-AGTGTCTCACTTCA	200
Qy	201	TACCGGGGAGCT	CGAGACTTCACTAGTGAATGACCCGAGCTTCTTTATGATG	260
Db	201	TACCGGGGATTTCTAT	CACTTCAATGCTCCACATGATATCCCATGATGTCTTTATATGATG	260
Qy	261	CCGTTTGGCCTGA	TATGATCTGACCAATGAAACAACAAATTTTGGCCCCCACTACAT	320
Db	261	CCATTTATCTGA	TAGAGAACTCACTGTGTAACAACAAGTATTTGGCGAACTTCA	320
Qy	321	AAAGAACCAT	TCAAGCAGCTCCAGGTGTGTGTGGAGAGCAAGCTGTTTAAAGTA	380
Db	321	AAGGAACTAT	TCAAGCAGCTCCATTAATGTGTGGAGAGCAAGCTACTTTAAATTA	380
Qy	381	TCTTTCAGAA	AGAGGTGCATGAAATTTGGCCCACTGAATGATAAAGCTGCTGCT	440
Db	381	GTCTTCAGAA	ATGAGGTGCATTGAAATTTGCCCATGTAATGTAAGTAAAGCTGCTGCT	440
Qy	441	GCTCCAGAGAA	TTTCCACTTGAAGTATAATTAAGTGTCCACACTTCAGACTGAC	500
Db	441	GCTGCCAGAGAT	TTTCCACTTGAAGTATAATTAAGTGTCCAGCTCTATAGGAATTTAT	500
Qy	501	ATPAAATAC	TGTCCTCAGAGGTGCTCTTCAACAAGCACTGTTCACGATATCCA	560
Db	501	GTAATTAAC	TGCGGAAGGGAAAT---ATATGCACTCCACAGATGCTTGTTCAG-----	549
Qy	561	ATTGTGATCT	TATGACCCCAACAACAGAGATATACAGTCCAAACAAGGGGAATATGAACT	620
Db	550	-TTATTTGCT	TATGAGCCCAACTCAGAGAAATATGAGCCCACTCCCGGA--TAGGAGC	606
Qy	621	CCACACAGAA	TATGAGCCCAACAAGGGGGAATATGAGCCCAACTATAGGAATA	676
Db	607	CCCACTGAGAA	TATGAGCCCAACAAGGTAGGAATAATAGAGTCCGCTGTAGGAATA	662

RESULT 4	
BG722815	775 bp mRNA linear EST 08-MAY-2001
LOCUS	602699527BP1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4827427 5',
DEFINITION	mRNA sequence.
ACCESSION	BG722815
VERSION	BG722815.1 GI:14002002
KEYWORDS	EST.
SOURCE	Homo sapiens (human)

REFERENCE 1 (bases 1 to 775)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.

Tissue Procurement: Miklos Palcikovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:

FEATURES	Location/Qualifiers
SOURCE	1. .775

```
/organism="Homo sapiens"  
/mol_type="mrna"  
/db_xref="taxon:9606"  
/clone="IMAGE:4827427"
```


FEATURES
SOURCE

Location/Qualifiers
1. 717
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4930521123"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"
/clone_lib="RIKEN full-length enriched, adult male testis"
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'. cDNA was prepared by using triethanol thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 5'
GAGAGAGAGAGCTCGGCAATTAATTCGAGTAAATTAATCCCCCCCCC -3'. cDNA was cloned into the XhoI and BamHI sites."

BASE COUNT 161 a 193 c 198 g 165 t

ORIGIN

Query Match 24.2%; Score 341.8; DB 14; Length 717;
Best Local Similarity 70.6%; Pred. No. 9.6e-67;
Matches 506; Conservative 0; Mismatches 192; Indels 19; Gaps 3;

1 CGGCAAGAGGGCGGCGAGAGGGGGCTT-GGGCAGAGATGCGAGTAAACCAAGCCACC 59
19 CGGCGGAGGGCGGCGGAG 78
60 GAGAGCCGCTGCGGGGCTCTATCCCTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 119
79 GTGAGCCGCGGTGGGGCGGCGAGTCCCTCATGTGTGAAGTCTCTTAAGAGAGAGAGAG 138
120 GTGAGCTCTGCTCTCTAG 179
139 GTGAGCT 198
180 GGAAGCTGTTCTCATCTATACCGGGTGTCTTCTGTGACTTCACTACTAGTATGAC 239
199 GGGGCTCTGTTCTACCTCATACCGGGTGTCTTCTGTGACTTCTCGGTGACAGACAGAT 258
240 CCCATGCTTCTTTTATGATGCGGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 299
259 CCCATGCTTCTTTTATGATGCGGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 318
300 ATTTTGGCCCAACATTAAG 359
319 ATCTTTGTGCAAACTACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 378
360 GGAAGAGCTGTTTAAAGTATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
379 GGTTCGTACTTAAATATGCTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 438
420 ATGTTAAAGTCTCTGCTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 479
439 ATGCGCAAGAGCTCTGCGGCTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 498
480 TTGCAACTTCAAG 539
499 ATGAGGCTCTAG 557
540 AACCTTTGTCAGATATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 599
558 ATATCC-----AAGTTGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600

QY 600 CAACAGAGGAGATATGAGATCTCCACAGAGAGATATGAGAGAGAGAGAGAGAGAGAGAG 659
DB 601 TCAACCTGAGATATGAG 660
QY 660 GCCCAACCTATGAGATATGAG 716
DB 661 GCCCGGCGCGGTGGGTATGATATGATATGATATGATATGATATGATATGATATGATATG 717

RESULT 6

BG701881

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

701 bp mRNA linear EST 07-MAY-2001
602683237F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4815849 5',
mRNA sequence.
BG701881.1 GI:13972665
EST.
Homo sapiens (human)
Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 701)
NIH-MGC <http://mgs.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Mikhail Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.llnl.gov>
Plate: L1AM10713 row: j column: 10
High quality sequence stop: 680.

FEATURES
SOURCE

Location/Qualifiers
1. 701
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4815849"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (GCGAG
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTT-3',
size-selected for average insert size 2.5 kb and
normalized to R07 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC library."

BASE COUNT

176 a 164 c 189 g 172 t

ORIGIN

Query Match 15.0%; Score 211.4; DB 10; Length 701;
Best Local Similarity 67.9%; Pred. No. 2.9e-37;
Matches 464; Conservative 0; Mismatches 191; Indels 28; Gaps 11;

QY 21 GGGGGCTGGGAGAGATGAG 80
DB 22 GAGGCCCAAGCAAGATGCGGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 80
QY 81 ATCCCTCTGGGAG 139
DB 81 ATCCCTTAACGAG 140
QY 140 GAAACAGATGATCTATCTTTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 199
DB 141 GCGATCAGAGAGCTCAATATGCTTTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 200

QY 200 ATACCGGAGTGTCTGTGAGCTTCAACCTTAAGTCANAGACCCAGTCT--TCTTTTANG 257
 DB 201 ATACCGGAGTGTCTGTGAGCTTCAACCTTAAGTCANAGACCCAGTCTCTTTTANG 260
 QY 258 ATACCGGAGTGTCTGTGAGCTTCAACCTTAAGTCANAGACCCAGTCTCTTTTANG 312
 DB 261 ATACCGGAGTGTCTGTGAGCTTCAACCTTAAGTCANAGACCCAGTCTCTTTTANG 320
 QY 313 ACTACCTTAAGGAGACCTTCAAGGAGCTTCAAGGAGCTTCAAGGAGCTTCAAGGAGCTT 372
 DB 321 CTTCGATTAAGGAGACCTTCAAGGAGCTTCAAGGAGCTTCAAGGAGCTTCAAGGAGCTT 380
 QY 373 TT---AGTTATCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 428
 DB 381 TTTATTAAGTGTCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 440
 QY 429 GCTGCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 487
 DB 441 GCGGGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 500
 QY 488 TTCAGAGCTGTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 547
 DB 501 TATGGGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 559
 QY 548 TCCAGATATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 607
 DB 560 TACAG-----TTATGTCTATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 609
 QY 608 GGAATATGAGAACTCCACAGAGAGATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 667
 DB 610 GGA--TACGAGGCCCACTGTGAGAGATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 666
 QY 668 TATGGATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 690
 DB 667 TGTGGATATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 689

RESULT 7
 BY098622 360 bp mRNA linear EST 07-DEC-2002
 LOCUS BY098622 RIKEN full-length enriched, pooled tissues, adult spleen,
 DEFINITION etc. Mus musculus cDNA clone K630123B09 5', mRNA sequence.
 ACCESSION BY098622
 VERSION BY098622.1 GI:26209239
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 360)
 Nkaido, I., Furuno, M., Kasukawa, T., Adachi, U., Bono, H., Kondo, S.,
 Okazaki, Y., Osato, N., Saito, R., Suzuki, H., Yamana, K., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gajdosi, I., Balderelli, R., Hill, D. P., Bulc, C., Hume, D. A.,
 Quakenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batilov, S.,
 Beitel, K. W., Blake, J. A., Bradt, D., Brats, V., Chotila, C., Coban,
 L. E., Cousins, S., Dalla, B., Dragani, T. A., Fletcher, C. P., Forrest,
 A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Glasi, C., Godzik, A.,
 Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J.,
 Jariab, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, K. M.,
 King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons,
 P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
 H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G.,
 Pesole, G., Petrovsky, N., Piliat, R., Pontius, J. U., Qi, D.,
 Ramchandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring,
 B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C. A., Setou,
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale,
 R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,
 Watanabe, Y., Welle, C., Wilming, L. G., Wymahaw-Boris, A., Yanagisawa,
 M., Yang, H., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
 Carinici, P., Hayatsu, N., Hirozane-Kishikawa, T., Kono, H., Nakamura,
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 COMMENT
 12466851
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Sphiro-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gs.c.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Aizawa, K., Akimura, T., Arakawa, T., Carinici, P., Fukuda, S., Hirozane,
 T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,
 Murata, M., Nakamura, M., Nomura, K., Numata, K., Ohno, M., Sakai, K.,
 Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami,
 M., Waki, K., Watanabe, A., Watanabe, M., and Hayashizaki, Y. Direct
 Submissions

Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multichannel sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

FEATURES

source
 1..360
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone_id="K630123B09"
 /clone_lib="RIKEN full-length enriched, pooled tissues,
 adult spleen, etc."
 /note="(dev stage=adult, tissue type=spleen, sex=male),
 (dev stage=adult, tissue type=kidney, sex=male),
 (dev stage=adult, tissue type=testis, sex=male),
 (dev stage=adult, tissue type=thymus, sex=male),
 (dev stage=adult, tissue type=colon, sex=male),
 (dev stage=adult, tissue type=heart, sex=male),
 (dev stage=adult, tissue type=stomach, sex=male),
 (dev stage=adult, tissue type=liver, sex=male),
 (dev stage=13 days embryo, tissue type=whole body, sex=mix
), (dev stage=14 days embryo, tissue type=whole body, sex=mix
), (dev stage=16 days embryo, tissue type=whole body, sex=mix
), (dev stage=17 days embryo, tissue type=whole body, sex=mix
), (dev stage=15 days pregnant, adult, tissue type=amion
 , sex=female), (dev stage=10 days neonate, tissue type=brin
 , sex=mix), (dev stage=10 days neonate, tissue type=thymus
 , sex=mix), (dev stage=10 days neonate, tissue type=heart
 , sex=mix)"

BASE COUNT

89 a 91 c 94 g 86 t

Query Match 14.0%; Score 197.8; DB 13; Length 360;
 Best Local Similarity 75.7%; Pred. No. 3.6e-34;
 Matches 258; Conservative 0; Mismatches 82; Indels 1; Gaps 1;

QY 1 CGGACAGAGGGGCGGAGAGGGGCTT-GGGAGAGTGGCAGTGAACGAGAGCCACC 59
 DB 20 CGGCGAGAGCGGCGGCGGAGAGAGAGCTTAAAGCAAGATGACAGAAACCAATACA 79
 QY 60 GAGAGCCGTCGGGGGCGCTCATCCCTTGCGCAAGTGTCTTGAAGAGTGTAGAT 119
 DB 80 GTGAGCCGCGCTTGGGCGGCGCATCTCATGTGAAGATCTCTTGAAGAAATGTTCAGAA 139
 QY 120 GTGAGCCTGTGCTTCTTCAAGAAACAGTGAATCTTATCTTTAAATGCGACAAAGAA 179
 DB 140 GTGAGCCTGTCTTCTTCCCGAGTACCGCAGGCTCCAACTCTTAACTGTGAACAAAGAG 199
 QY 180 GGAACGTTGTTTCTCATTTACACCGGTGTCTTCTGTGACTTCACTTAATGCAATGAC 239
 DB 200 GGGGCTGTGTTTCTGACCTCAATACCGGTGTATTTTGTGACTTGTGCGAGACACGAT 259
 QY 240 CCATGCTTCTTTATGATGACCGCTTGCGTGTGATGATGATGATGATGATGATGATGAT 299
 DB 260 CCATGCTTCTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 319
 QY 300 ATTTTGGCCCCCAATTAATTAAGAAACCAATTCAGGACG 340
 DB 320 ATCTTGTGCAAACTAATTAAGGAGCAATTCAGGACG 360

RESULT 8
 BUI07273 932 bp mRNA linear EST 25-NOV-2002
 LOCUS 60311954F1 CSBQCHL12 Gallus gallus CDNA clone CHEST60K13 5', mRNA
 DEFINITION sequence.

ACCESSION BUI07273
 VERSION BUI07273.1 GI:25309754
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus

REFERENCE 1 (bases 1 to 932)
 Boardman,P.R., Sans-Beguerro,J., Overton,I.M., Burt,D.W., Bosch,B.,
 Pong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 A Comprehensive Collection of Chicken CDNA
 Curr. Biol. 12 (22), 1965-1969 (2002)

TITLE JOURNAL MEDLINE PUBMED
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 10D, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source
 1..932
 /location/Qualifiers
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST60K13"
 /dev_stage="36"
 /lab_host="DH10B"
 /note="Organ: heart; Vector: pBluescript II KS(+); Site 1:
 EcoRI, Site 2: NotI; Modification of pBluescript II KS(+)
 [Stratagene] vector to accommodate cDNA produced with the
 T-primed protocol (Construction of uni-directionally
 cloned cDNA libraries from messenger RNA for improved 3'
 end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
 ,624). Out pBluescript II KS(+) with NotI and EcoRI.
 ligase in double stranded adaptor containing BglI and
 BamHI sites [5'ggccgcgcgcagcccgatccgaataaag]

BASE COUNT 215 a 273 c 233 g 211 t
 ORIGIN [5'aattccttttttcgataccggggctgcagc]

Query Match 13.0%; Score 183.6; DB 13; Length 932;
 Best Local Similarity 59.5%; Pred. No. 5,4e-31;
 Matches 350; Conservative 0; Mismatches 229; Indels 9; Gaps 2;

QY 42 GTGAACCAAGCCACACCGAGAGCGTGTGGGGCCCTCATCCCTCTGGCGAAAGTGC 101
 DB 21 GCGCTCAATTAAGGACACCTCGAAGAGGCGGTGTGTGTCTCCCAATGCCAGACATT 80
 QY 102 TTGAACAGTGTGAGATGTGACCTCTGCTTCTTCAAGAAACAGTGAATCTATCTC 161
 DB 81 CTGACCATGTAAAGATGTGAGCTCTCTTCAAGACATGACGCGCAAGCTTGAAGCC 140
 QY 162 TTTAATGCAAAAGAAAGAAAGCTTTTCTCACTTCAATACGGGTGTCTTGTGACT 221
 DB 141 TTCAAAAGGACCAAGAAAGAAAGTGTGTATCTCAACCCCATACAGATGATCTTGTGTC- 199
 QY 222 TCACACTTATGTGAATGACCCCATGCTTCTTTTATGATGCCGTTGGCCGTATGATGAC 281
 DB 200 -----AAAGGCAAGGATCTTATGCTCTTTATATATCCGTTTATTTGTGAAGAG 254
 QY 282 TGCAACATTGAACAAACCAATTTTGGCCCACTACATTAAGAAACCAATTCAGGAGCT 341
 DB 255 TGCTCTATTGAGAGAGCTCTTTCTCTGTCTATTAATTAATTAAGACAGATTCAGCTGAG 314
 QY 342 CGAGTGTGTGCTGTGGAAGACAAAGCTTTTAACTTAATCTTCAGAAAGAGGTGCC 401
 DB 315 GCAGAGAGTGTGCGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 374
 QY 402 ATGGAATTTTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 461
 DB 375 ATGGAATTTTGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 434
 QY 462 GGAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 521
 DB 435 CAGCCCCCTGTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 494
 QY 522 GCAATGTGCTCTTCAAGACACC--TTGTCCAGCATTCATTAATGATGATGATGATGATGAT 578
 DB 495 CAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 554
 QY 579 CACACCAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 626
 DB 555 GCTCCAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 602

RESULT 9
 BUI61805 814 bp mRNA linear EST 21-OCT-2002
 LOCUS AGENCOURT 10617160 NIH_MGC_169 Mus musculus CDNA clone
 DEFINITION IMAGE:6742447 5', mRNA sequence.

ACCESSION BUI61805
 VERSION BUI61805.1 GI:24191377
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 814)
 NIH-MGC http://mgc.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgabs-remail.nih.gov

Tissue Procurement: Dr. Jonathan Kuo, NIH
 CDNA Library Preparation: Michael Brownstein Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.jnl.gov
 Plate: L1C3080 row: 1 column: 06
 High quality sequence stop: 504.
 Location/Qualifiers

FEATURES

Source

1..814
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6742447"
 /lab_host="DH10B (T1-phage-resistant)"
 /clone_1lb="N1H_MGC_169"
 /note="Organ: Testicles; Vector: pDNR-LIB; Site 1: 8fil (ggccatcagcc); Site 2: 8fil (ggccgcgcgcgc); cDNA made by oligo-dt priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGGCGATACGACGATGCGCAGATGCGCGG-3' and
 5'-ATCTAGAGCGCGAGCGCGCGCAGATG-3' (30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in the laboratory of M. Brownstein (NIMH, NIH). Note: this is a N1H_MGC Library."
 BASE COUNT 207 a 238 c 177 g 190 t 2 others
 ORIGIN

Query Match 12.9%; Score 182.2; DB 13; Length 814;
 Best Local Similarity 65.1%; Pred. No. 1.1e-30;
 Matches 304; Conservative 0; Mismatches 153; Indels 10; Gaps 2;
 615 GGAATCCACAGAAAGATATGAGCCCAAGCGGAGATATGAGCCCACTTATGGA 674
 3 GGCTCCCTCTCTCTATATGATGATACCCCAATGGGTATGAGTTCACTCTCGGA 62
 675 TATGAGCCCGCTGTGAGATATGAGTCCCACTGGGGATATGAGTCCCACTGGG 734
 63 TATGAGCTCCACCTGTGAGATACGATCCCACTCTGATATGAGACCCCACTATG 122
 725 GATATGAGTCCCACTGGGGATATGAGCCCACTGGGGATATGAGTCCCACT 794
 123 GATATGAGCCCACTCTGATATGAGATCCCACTCTGATATGAGTCCCACTCT 182
 795 GGGGATATGAGTCCCACTGGGGATATGAGCCCACTGAGATATGAGCCCA 854
 183 CCCAGATATGAGCCCGCTTATGAGGATATGAGTCCCTTCAAGGGGATAGTCAT 242
 855 CCAAGTGAATGAAGCCCTTACCCCTGATATGAGTCCCTGCTGAGAAATGAGT 914
 243 CCTCTGATCAAGGGCCAGCTCTGATGACAGAGAGCTCACTGCTGAGTCAAGCA 302
 915 GCTCTCAGATCTATGACAGCTCAGCA-----GAGACTTCTTCCCACTACC 965
 303 GGCATCCCATGCTGTGAGATGCAATCTGATTCAGAGCATTTTCTCTTACC 362
 966 TCATCTTCTAGTCCATTTACCACTTCTGAGAGTTAACTTGAAGATCAGCAAGC- 1024
 363 TCGTCTCAGAGTCACTCTCCGCTTAAAGATTAACCTTGAAGTTTCAACAGCA 422
 1025 AAAAGGCACTTAAACTGAGTCAAGTGAAGAGAGAGCCAGT 1071
 423 AAAAGGCACTTAACTGAGTCAAGATTAACCCGGAACCTTAAAGT 469

RESULT 10
 BUI03740 560 bp mRNA linear EST 25-NOV-2002
 LOCUS 60305979P1 CSBQCHL01 Gallus gallus CDNA clone CHS8T23n20 5', mRNA
 DEFINITION
 sequence.
 ACCESSION BUI03740
 VERSION BUI03740.1 GI:25304305
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Bacteria; Metazoa; Chordata; Craniata; Vertebrata; Neoteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

REFERENCE

1 (bases 1 to 560)
 Boardman, P.B., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, B.,
 Pong, W.T., Rickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.V.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 JOURNAL MEDLINE 22335534
 PUBMED 12445392

FEATURES

Source

1..560
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHS8T23n20"
 /tissue_type="whole embryo"
 /dev_stage="20-21"
 /lab_host="DH10B"
 /clone_1lb="CSBQCHL01"
 /note="Organ: whole embryo; Vector: pBluescript II KS(+);
 Site 1: EcoRI; Site 2: NotI. Modification of pBluescript
 II KS(+) [Stratagene] vector to accommodate cDNA produced
 with the T-primed protocol (construction of
 uni-directionally cloned cDNA libraries from messenger RNA
 for improved 3' end DNA sequencing by Glenn Fu, et al.
 U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with
 NotI and EcoRI. Ligate in double stranded adaptor
 containing BglI and BamHI sites
 [5'-ggcgctcgacggccggatcggagaaagc] 5'-attctcttttcggaccggcgccgcgc]
 [5'-attctcttttcggaccggcgccgcgc]
 BASE COUNT 131 a 142 c 159 g 128 t
 ORIGIN

Query Match 12.7%; Score 179.4; DB 13; Length 560;
 Best Local Similarity 66.1%; Pred. No. 5e-30;
 Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;
 42 GTGAACAGAGCAACCGAGAGCGGTGTGAGGCGCTTATCCCTCTGCGAAAGTTC 101
 42 GCGCTCATTAAGCACTCCAGAGAGAGCGGTGTGAGTCCCAATGCCAATGCAATT 101
 102 TTGAAGCATGTGAGATGTGACTCTGCTTCTCAAGAAACAGTGAATCTATCTC 161
 102 CTCAAGCATGTAAAGATGTGAGCTCTCTTCAAGTGAATGACGGCAAGCTTGAAGCC 161
 162 TTTAATGCAAGAAAGAAAGAAAGCTTTTCTCACTTATCCCGGTGTCTTCTGACT 221
 162 TTCAAGGACCAAGAAAGAAAGTGTATCTACCCCAATCAAGATATCTTCTGTCTC 220
 222 TCACACTTATGATGACCCCATGCTTCTTATGATGAGGCTTGTGCTGATGATGAC 281
 221 -----AAAGGCAAGATCTATGCTGTCTTATGATGAGGCTTATGATGATGAGG 275
 282 TGACCATTTGAACCAATTTTTCGCCCAATGATTAAGAAACATTCAGGCACT 341
 276 TGCTTATTTGAGAGCGCTCTTCTTCTCTATATATCAACAAAGGCAATTCAGGCTGAG 335
 342 CCAAGTGTGCTGAGAGGAGCAAGCTTTTAAATGATATCTTCAAGAAAGAGAGTCC 401
 336 GCAAGAGGTGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 395
 402 ATCGAATTTGCCCACTGATGATTAAGAGTCTCTGTGCTGCTGCAAGAGAAATTCACCT 460
 396 ATCGAGTTTGAACAGCTGATGATTAAGAGTCTCTGTGCTGCTGCAAGAGAGGAGTCTCT 454

```

RESULT 11
LOCUS      BU284977
DEFINITION BU284977 700 bp mRNA linear EST 27-NOV-2002
            603603379P1 CSEQCEN54 Gallus gallus cDNA clone CHEST581n20 5', mRNA
            sequence.
ACCESSION  BU284977
VERSION    BU284977.1 GI:25734433
KEYWORDS   EST.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 700)
            Boardman,P.B., Sanz-Bzquez,J., Overton,I.M., Burt,D.W., Bosch,E.,
            Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
            A Comprehensive Collection of Chicken CDNA
            Curr. Biol. 12 (22), 1965-1969 (2002)
JOURNAL    22335534
MEDLINE    12445392
PUBMED
COMMENT    Contact: Simon Hubbard
            Department of Biomolecular Sciences
            University of Manchester Institute of Science and Technology (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612360409
            Email: Simon.Hubbard@umist.ac.uk.
            Location/Qualifiers
            1..700
               /organism="Gallus gallus"
               /mol_type="mRNA"
               /strain="Compton line 151"
               /db_xref="taxon:9031"
               /clone="CHEST581n20"
               /sex="Female"
               /tissue_type="not cerebrum or cerebellum"
               /dev_stage="adult"
               /lab_host="DH10B"
               /clone_id="CSEQCEN54"
            /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
            EcoRI; Site_2: NotI; This normalized library was
            constructed from 1 million independent clones. cDNA
            synthesis was initiated using an oligo(dT) primer, using
            methylated C in the first strand synthesis reaction.
            Following this first strand reaction, double-stranded cDNA
            was bluntend, ligated to NotI adapter, digested with EcoRI
            , size-selected, and cloned into the NotI and EcoRI
            compatible sites of a custom modified MCS of the
            pBluescript (KS+) vector. The library was normalized in 2
            rounds using conditions adapted from Soares et al., PNAS
            (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
            (1996): 791, except that a significantly longer
            reannealing hybridization was used."

BASE COUNT      162 a      196 c      178 g      164 t
ORIGIN
Query Match      12.7%; Score 179.4; DB 13; Length 700;
Best Local Similarity 66.1%; Pred. No. 4.9e-30;
Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY      42 GTGAACCAAGCCACACCGAGCGCGTGTGGGCGCTCATCCCTCTGCGAAAGTGC 101
DB      6 GCGCTCAATAGGCACCACTCGAAGAAAGCGGTGTCGTCGCCCAATGCCAGAGCAT 65
QY      102 TTGAAGAGGTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 161
DB      66 CTCAAGCAATGTAAGATGTAAGATGTAAGATGTAAGATGTAAGATGTAAGATGTAAG 125
QY      162 TTTAATGGCAAAAGAAAGAAAGTGTCTTCTCACTTCAATCCGGGTGCTTCTGTGACT 221
DB      126 TTCAAGGCAACCAAGAGGAGATGCTATCTCAACCCCATACAGATGATCTTGTGTGTC- 184

```

```

QY      222 TCACACTTAGTCGAAGACCCAGCTCTTCTTTAAGATGCCGTTGGCCGTGATAGTAC 281
DB      185 -----AAAGGCAAGATCTCTATGCTGTCTTTATATAGCGCTTTATTTGTGAAGG 239
QY      282 TGCACCATTTGAAACCAATTTTGGCCCACTAATTAAGAAACCAATTCAGGAGCT 341
DB      240 TGCCTATTGAGAGAGCTCTTTCTGCTGATTAATCAAGAGAGATTCAGGCTGAG 299
QY      342 CCAAGTGTGCTGCGAAGAGCAAGCTGTTTAACTTATCTTCAAGAAAGAGTGC 401
DB      300 GCAGAGAGTGTGCTGCGAAGAGCGAGAGAGTAACTGACTTCAACAGCGAGAGG 359
QY      402 ATGCAATTTGCCCACTGATGTAAGATGCTGCTGCTGCTGCCAGAGATTCAC 460
DB      360 ATGCAATTTGCAAGCTGATGTAAGATGCTGCTGCTGCTGCCAGAGATTCAC 418

RESULT 12
LOCUS      AJ452591
DEFINITION AJ452591 723 bp mRNA linear EST 22-APR-2002
            AJ452591 riken1 Gallus gallus cDNA clone 31her1, mRNA sequence.
ACCESSION  AJ452591
VERSION    AJ452591.1 GI:20262687
KEYWORDS   EST.
SOURCE     Gallus gallus (chicken)
ORGANISM   Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
            1 (bases 1 to 723)
REFERENCE  Buerstedde,J.M.
            Gallus gallus bursal lymphocyte EST
            Unpublished
            Contact: Buerstedde JM
            Cellular Immunology
            Heinrich-Pette-Institute
            Martinistr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
            Location/Qualifiers
            1..723
               /organism="Gallus gallus"
               /mol_type="mRNA"
               /db_xref="taxon:9031"
               /clone="31her1"
               /cell_type="bursal lymphocyte"
               /dev_stage="2-3 weeks old"
               /clone_id="riken1"
               /note="CB indred strain"

BASE COUNT      161 a      205 c      181 g      172 t      4 others
ORIGIN
Query Match      12.7%; Score 179.4; DB 9; Length 723;
Best Local Similarity 66.1%; Pred. No. 4.9e-30;
Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY      42 GTGAACCAAGCCACACCGAGCGCGTGTGGGCGCTCATCCCTCTGCGAAAGTGC 101
DB      15 GCGCTCAATAGGCACCACTCGAAGAAAGCGGTGTCGTCGCCCAATGCCAGAGCAT 74
QY      102 TTGAAGAGGTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 161
DB      75 CTCAAGCAATGTAAGATGTAAGATGTAAGATGTAAGATGTAAGATGTAAGATGTAAG 134
QY      162 TTTAATGGCAAAAGAAAGAAAGTGTCTTCTCACTTCAATCCGGGTGCTTCTGTGACT 221
DB      135 TTCAAGGCAACCAAGAGGAGATGCTATCTCAACCCCATACAGATGATCTTGTGTGTC- 193
QY      222 TCACACTTAGTCGAAGACCCAGCTCTTCTTTAAGATGCCGTTGGCTGATGATGAC 281
DB      194 -----AAAGGCAAGATCTCTATGCTGTCTTTATGATGCGTTTATTTGTGAAGG 248
QY      282 TGACCATTTGAACAACAATTTTGGCCCACTACATTAAAGAACATTTCAGGAGCT 341

```


Db	249	TGCTCTATTGAGCAACCTCTTTCTCTGCTAATATCAACAAAGACAATTCAGGCTAG	308
Qy	342	CGAGTGTGCTGTGGAAAGACACAGCTGTTTAAAGTATCTTCAGAAAGAGGTGCC	401
Db	309	GCAGAGGTGCTGTGGAAAGGCGACAGGAACGTTAAATGACTTTCAACAGCGAGAGACC	368
Qy	402	ATCGAATTTGCCCACTGATAGTAAAGAGTGTGCTGTGCTGCTCCAGAGAAATTCCTACT	460
Db	369	ATCGAGTTTGACACACTGATATGTTCAAAAGCTGCTCTAGTGTCTTCAGTGGAGTTCTCT	427
RESULT 13			
BI390687			
LOCUS			
DEFINITION			
pspin.pk001.p22 Normalized Chicken Ptitutary/Hypothalamus/Pineal			
Library Gallus gallus cDNA clone Pspin.pk001.p22 5' similar to			
gbl/AAD10951.1 (U79458) WW domain binding protein-2 (Homo sapiens),			
mRNA sequence.			
ACCESSION			
BI390687			
VERSION			
BI390687.1			
KEYWORDS			
EST.			
SOURCE			
ORGANISM			
Gallus gallus (chicken)			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;			
Phasianinae; Gallus.			
1 (bases 1 to 568)			
Porter,T.R. and Cogburn,L.A.			
ESTs from Normalized Chicken Ptitutary/Hypothalamus/Pineal cDNA			
Library USDA/IRPS Animal Genome Project			
Unpublished			
JOURNAL			
COMMENT			
Contact: Larry A. Cogburn			
University of Delaware			
Townsend Hall, Newark, DE 19717, USA			
Tel: 302-831-1335			
Fax: 302-831-2822			
Email: cogburn@udel.edu, www.chickest.udel.edu.			
FEATURES			
source			
1..568			
location/Qualifiers			
/organism="Gallus gallus"			
/mol_type="mRNA"			
/strain="Commercial broiler chicken"			
/db_xref="taxon:9031"			
/clone="Pspin.pk001.p22"			
/sex="Male and Female"			
/tissue_type="Ptitutary Gland/Hypothalamus/Pineal Gland"			
/dev_stage="Embryonic (d12,d14,d19); post-hatch (w1,w3,w5			
'w7,w9)"			
/lab_host="B. Coli EMDH10B"			
/clone_lib="Normalized Chicken			
Ptitutary/Hypothalamus/Pineal Library"			
/note="Vector: pCMVSPORT6; Library made from equivalent			
pools of total RNA isolated from each tissue at different			
ages. Single pass sequencing from 5'-end"			
BASE COUNT			
128 a 146 c 159 g 130 t 5 others			
ORIGIN			
Query Match			
Best Local Similarity 12.6%; Score 177.8; DB 12; Length 568;			
Matches 276; Conservative 0; Mismatches 137; Indels 6; Gaps 14			
Qy	42	GTGAACCGAGACACACCGAGACCGGTGTGGGCGCCCTCATGCCCTCTGGGAAAGTGC	101
Db	26	GCGCTCATTTAGGCACCACTCGAAGAGACGGGTGTGCTGCTCCCAATGCCAGAGCAATT	85
Qy	102	TTGAAGCAGTGTGAGATGTGACCTCTGCTTCTTACAGAAACAGTGAATCTTATCTC	161
Db	86	CTCAAGCAATGTAAAGATGTGGAGCTCTCTTCACTGACATGACGGCGACGCTTGAAGCC	145
Qy	162	TTTAATGGCAAAAGAAAGAAACGTTTGTTCCTTCACTTATACCGGTGTCTTGTACT	221
Db	146	TTCAAGGACACCAAGAGGAATGTCTGTATCTCAACCCATTCAGATATCTTGTGTTC	204

OY		222	TGACACTTAAGTCANATNAGACCCTTCTTTTANATNGATCCGTGGCTTGATGATGAC	281
Dd		205	-----AAAGGCCAAGGATCCATACCTGTCTTTTATGATGCATTATTTGGTAGAAGGG	259
OY		282	TGCACCAATTTGAACAACCAATTTTGACCCCACTACATTAAGAAGCAATTCAGGCACT	341
Dd		260	TGCTCTATATAGACAGCCTCTTTTCTCTGTCTATTTACATCAAAGACAGATTCAGGCTGAG	319
OY		342	CCAGGTGTGTGCTGGAGAGCAAGCTGTTTTTAAGTTATCTTCAGAAAAGAGGTGCC	401
Dd		320	GCAAGAGGTGTGCTGGAGAGGCAAGGAAACCTTTAACTGACTTTCACACAGGAGAGAGCC	379
OY		402	ATCGAATTTGGCCCACTGATGTGTAAAGCTGCTCTGTCTGTGCACAGGAATTCCTACT	460
Dd		380	ATCGAGTTTGGACAGCTGATGTGTTCAAAGCTGCTCTGTGTGCTTCAGTGGAGTTCTCT	438
 RESULT 14 BM490729				
LOCUS		BM490729	636 bp	mRNA linear EST 07-FEB-2002
DEFINITION		pgp2n.pk004.e6 Normalized Chicken Pituitary/Hypothalamus/Pineal Library (pgp2n) Gallus gallus cDNA clone pgp2n.pk004.e6 5' similar to gblAAD10951.1 (U79458) WM domain binding protein-2 [Homo sapiens], mRNA sequence.		
ACCESSION		BM490729		
VERSION		BM490729.1	GI:18611660	
KEYWORDS		EST.		
SOURCE		Gallus gallus (chicken)		
ORGANISM		Gallus gallus		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.		
REFERENCE		Porter,T.B. and Cogburn,T.A.		
AUTHORS		1 (bases 1 to 636)		
TIITLE		ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal CDNA library, USDA/IRAPs Animal Genome Project		
JOURNAL		Unpublished		
COMMENT		Contact: Larry A. Cogburn		
		University of Delaware		
		Townsend Hall, Newark, DE 19717, USA		
		Tel: 302-831-1335		
		Fax: 302-831-2822		
		Email: cogburn@udel.edu, www.chickest.udel.edu.		
FEATURES		Location/Qualifiers		
Source		1..636		
		/organism="Gallus gallus"		
		/mol_type="mRNA"		
		/strain="Commercial broiler chickens"		
		/db_xref="taxon:9031"		
		/clone="pgp2n.pk004.e6"		
		/sex="Male and Female"		
		/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"		
		/dev_stage="Embryonic (dl2,d14,d19), post-hatch (1,3,5,7,9 weeks)"		
		/lab_host="B. coli BMDH109"		
		/clone_1lb="Normalized Chicken Pituitary/Hypothalamus/Pineal Library (pgp2n)"		
		/note="Vector: pCMVSPORT6; Library made from equivalent pools of total RNA isolated from each tissue at different ages. Single pass sequencing from 5'-end"		
BASE COUNT		143 a	174 c	146 t
ORIGIN				
Query Match		12.6%	Score 177.8;	DB 12; Length 636;
Best Local Similarity		65.9%;	Pred. No. 1.1e-29;	
Matches 276;		Conservative 0;	Mismatches 137;	Indels 6; Gaps 1;
OY		42	GTEAACAGAGCCACACACGAGAGCGCGTGTGGGCCCCCATCCCTCTGTGGAGAAATGTC	101
Dd		26	GCGCTCAATAGGCACTGAGAGAGAGCGGTGTGTGCTGCTCCCAATATGCCAAGACAT	85
OY		102	TTTGAGCAGTGATGAGATGTGACCTTGTCTTCTTACAGAAACGATGAAATCTTATCTC	161


```

Db      86 CTCAAGCAATGTAAAGATGTGAGCTCTCTTCAAGTACATGAGGCGAAGCTTGAGGCC 145
Qy      162 TTTTAATGGCAAAAGAAAGAAAGCTGTTTCTCACTTCATACCGGGTGTCTTGACT 221
Db      146 TTCAAAAGCACCAAGAAAGGAAATCTGTATCTCAACCCATACAGAAATCTTGTC- 204
Qy      222 TCACACTTAATGCAATGACCCCATGCTTCTTTTATGATGCGGTGGCGTGAAGTAC 281
Db      205 -----AAAGGCAAGATCTTATGCTGTCTTTTATGATGCAATTTATTTGGTAAAGGG 259
Qy      282 TGACACATTGAAACCAATTTTTCGCCCACTACATTAAGAACCATTCAGGAGCT 341
Db      260 TGCTCTATGAGAGCGCTCTTTCTGTCTAATTACATCAAGAGACAGATTCAAGGCTGAG 319
Qy      342 CCAGGTGTGCTGGGAAAGCAAGCTGTTTAAAGTTATCTTCAGAGAAAGAGTGGC 401
Db      320 GCAGAGGTGTGCTGGGAAAGGACAGGAAAGCTTTAACTGATCTTTCACAGGAGGAGGCC 379
Qy      402 ATCGAATTTGCCCACTGATGTAAAGCTGCTCTGCTGTCAGAGAGAAATTCGACT 460
Db      380 ATCGAGTTTGACAGCTGATGTTCAAAGCTGCTCTGAGTCTTCAAGTGAAGTCTCT 438

```

RESULT 15

```

AJ454950      772 bp      mRNA      linear      EST 22-APR-2002
LOCUS      AJ454950      riken1 Gallus gallus cDNA clone 4c7r1, mRNA sequence.
ACCESSION      AJ454950
VERSION      AJ454950.1 GI:20265046
KEYWORDS      EST.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus

```

```

REFERENCE      Buehrstedde J.M.
AUTHORS      Buehrstedde J.M.
TITLE      Gallus gallus bursal lymphocyte EST
JOURNAL      Unpublished
COMMENT      Contact: Buehrstedde JM
              Cellular Immunology
              Heinrich-Pette-Institute
              Martinistr. 52, 20251 Hamburg, Germany
              Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
              Location/Qualifiers

```

FEATURES

```

source
1..772
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="4c7r1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/clone_id="riken1"
/note="CB inbred strain"

```

```

BASE COUNT      176 a      154 c      210 g      232 t
ORIGIN
Query Match      11.8%; Score 167.4; DB 9; Length 772;
Best Local Similarity 65.1%; Pred. No. 2.5e-27;
Matches 265; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

```

```

Qy      42 GTGAACCAAGCCACACCGAGAGCGGTGTGGGCGCTCATCCCTCTGTGGGAAAGTGC 101
Db      23 GCGCTCAATAGGCACTCGAAGAAAGCGGTGTGTGTGCTCCCAATGCCGAGAGCAAT 82
Qy      102 TTGAAGAGTGTGAGATGTGACCTGTCTTCTACAGAAACAGTGAATCTATCTC 161
Db      83 CTCAGCAATGTAAAGATGTGAGCTCTCTTCAGTACATGATCGGCAAGCTTGAGGCC 142
Qy      162 TTTTAATGGCAAAAGAAAGAAAGCTGTTTCTCACTTCATACCGGGTGTCTTGACT 221
Db      143 TTCAAAAGCACCAAGAAAGGAAATCTGTATCTCAACCCATACAGAAATCTTGTC- 201

```

```

Qy      222 TCACACTTAATGCAATGACCCCATGCTTCTTTTATGATGCGGTGGCGTGAAGTAC 281
Db      202 -----AAAGGCAAGATCTTATGCTGTCTTTTATGATGCGGTGGTGAAGGG 256
Qy      282 TGACACATTGAAACCAATTTTTCGCCCACTACATTAAGAACCATTCAGGAGCT 341
Db      257 TGCTCTATGAGAGCGCTCTTTCTGTCTAATTACATCAAGAGACAGATTCAGGCTGAG 316
Qy      342 CCAGGTGTGCTGGGAAAGCAAGCTGTTTAAAGTTATCTTCAGAGAAAGAGTGGC 401
Db      317 GCAGAGGTGTGCTGGGAAAGGACAGGAAAGCTTTAACTGATCTTTCACAGGAGGAGGCC 376
Qy      402 ATCGAATTTGCCCACTGATGTAAAGCTGCTCTGCTGTCAGAGAGAAATTCGACT 448
Db      377 ATCGAGTTTGACAGCTGATGTTCAAAGCTGCTCTGAGTCTTCAAGTGAAG 423

```

Search completed: December 16, 2003, 17:31:37
Job time : 2917.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 05:59:34 ; Search time 339.84 Seconds

(without alignments)
10999.596 Million cell updates/sec

Title: US-09-864-291-4_COPY_36_933

Perfect score: 898
Sequence: 1 atggcagctgacacagagccca.....tgccttcacagatcatga 898Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sbs:*
12: gb_ey:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_dat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sbs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_ey:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	898	100.0	1413 6	AX359658
2	524.4	58.4	2267 9	BC022546
3	519.2	57.8	2266 9	BC022549
4	374.4	41.7	1001 6	AX359665
5	237	26.4	179222 2	BX296515
6	237	26.4	222469 2	BX470149
7	223.2	24.9	129624 2	BX296540
8	220	24.5	220895 2	HS250D10
9	156.8	17.5	168425 2	AC113593
10	156.8	17.5	224086 10	AC104325
11	151.8	16.9	253149 2	AC107527
12	151.8	16.9	270171 2	AC132969
13	128.4	14.3	793 10	AF499026
14	128.4	14.3	1802 10	MF040826
15	122.8	13.7	222469 2	BX470149
16	122	13.6	1691 9	HSU79458
17	122	13.6	1820 9	BC007452
18	122	13.6	1885 6	AR220846
19	122	13.6	1912 9	BC010616
20	121.6	13.5	2170 9	AB072784
21	120.4	13.4	1835 9	AK057881
22	116	12.9	1205 5	BC053232
23	116	12.9	95108 2	AC127818
24	110.2	12.3	106438 9	AL365445
25	106.4	11.8	2065 14	AF410847
26	106.4	11.8	148418 9	AC020698
27	103.8	11.6	454 6	BD027759
28	102.8	11.4	246814 2	AC103482
29	100	11.1	235381 2	AC103031
30	96	10.7	247947 2	AC098163
31	95.8	10.7	110000 2	AC118467_0
32	93.8	10.4	216292 2	AC100927
33	91.6	10.2	239972 2	AC095835
34	91.6	10.2	241519 2	AC097689
35	91.6	10.2	250902 2	AC131965
36	91.6	10.2	251746 2	AC131411
37	91.4	10.2	208880 2	AC128103
38	90.6	10.1	230569 2	AC108250
39	90.6	10.1	238377 2	AC121702
40	90.4	10.1	210680 2	AC123724
41	90.4	10.1	212334 2	AC023070
42	90.4	10.1	315390 2	AC124097
43	90	10.0	262967 2	AC110392
44	89.8	10.0	261914 2	AC122581
45	88.8	9.9	88716 2	AC098644

ALIGNMENTS

RESULT 1
AX359658
LOCUS AX359658 1413 bp DNA
DEFINITION Sequence 4 from Patent WO0190185.
ACCESSION AX359658
VERSION AX359658.1 GI:18675409
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
unclassified.

REFERENCES

1
Okamoto, R. and Sutoy, P.
P32 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses
thereof
Patent: WO 0190185-A 4 29-NOV-2001;
JOURNAL

QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES
UNIVERSITY (US)
Location/Qualifiers
1. 1413
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="mammalia"
36. 977
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD23336.1"
/db_xref="GI:18675410"
/translation="MAVNSHTSRKALIPSGESVILKQCEVDLCFLQKPEVSYLEN
GRKGTFLTSYVVPVSHLVNDPMLSPFMPQGLSDCTIROPFAPNPYIKGTIOAA
PGGMEGQAVFKLSFRKGAIPRAQIAAAGAPIGASVNYPDLSGLYITVP
GAIVSGQFCPAYPIYIGPPPGYVVGKXGKIPBGGAAPGGYGAAPPMYGAAP
VGVGVPBGYGVPPGGYGVPPGGYGAAPGGYGVPPGGYGAAPPGYGAAPPA
NEALPAYEAPSGNTASHSRNTAQDSTLPTTSSS"

BASE COUNT 377 a 363 c 369 g 304 t

Query Match 100.0%; Score 898; DB 6; Length 1413;
Best Local Similarity 100.0%; Pred. No. 2.4e-214;
Matches 898; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAATGAAACCAAGCAACCAAGCAAGCCGCTGTGGGCGCTCATCCCTCTGGGAA 60
DB 36 ATGCAATGAAACCAAGCAACCAAGCAAGCCGCTGTGGGCGCTCATCCCTCTGGGAA 95
QY 61 AGTGTCTTGAAGAGTGTGAAGATGTGAAGCTCTGCTTCTTAAGAAACCAAGTATCC 120
DB 96 AGTGTCTTGAAGAGTGTGAAGATGTGAAGCTCTGCTTCTTAAGAAACCAAGTATCC 155
QY 121 TATCTCTTAAAGCAAAAGAAAGAAAGCTGTCTCACTTCAATCCGGGTGTCTTC 180
DB 156 TATCTCTTAAAGCAAAAGAAAGAAAGCTGTCTCACTTCAATCCGGGTGTCTTC 215
QY 181 GTGACTTCACTTAAAGTAAAGCAAGCCGCTTCTTAAAGTAAAGCCGCTGTGAAG 240
DB 216 GTGACTTCACTTAAAGTAAAGCAAGCCGCTTCTTAAAGTAAAGCCGCTGTGAAG 275
QY 241 AGTACTGCAACATTTGATTAAGCAACCAATTTTGGCCCACTAATTAAAGAACCATTCAG 300
DB 276 AGTACTGCAACATTTGATTAAGCAACCAATTTTGGCCCACTAATTAAAGAACCATTCAG 335
QY 301 GCAAGCTCAGAGTGTGTGTGGAAGCAAGCTGTTTAAAGTAAAGTAAAGTAAAGTAAAG 360
DB 336 GCAAGCTCAGAGTGTGTGTGGAAGCAAGCTGTTTAAAGTAAAGTAAAGTAAAGTAAAG 395
QY 361 GGTGCAATGAAATTTGCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 420
DB 396 GGTGCAATGAAATTTGCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 455
QY 421 CCACTTGAAGTAAATTAAGTGTGCAACTTCAAGACTGTATCAATTAATTAATTAATTAATTA 480
DB 456 CCACTTGAAGTAAATTAAGTGTGCAACTTCAAGACTGTATCAATTAATTAATTAATTAATTA 515
QY 481 GGGGCTGCAAGTGTCTCTCAAGCAAGCTTGTCCCAAGTAAAGTAAAGTAAAGTAAAGTAAAG 540
DB 516 GGGGCTGCAAGTGTCTCTCAAGCAAGCTTGTCCCAAGTAAAGTAAAGTAAAGTAAAGTAAAG 575
QY 541 CCCCACCAACCAAGATTAAGTCAAGCAAGGGAATATGAATCTCAACCAAGAGATAT 600
DB 576 CCCCACCAACCAAGATTAAGTCAAGCAAGGGAATATGAATCTCAACCAAGAGATAT 635
QY 601 GAGGCCCAACCAAGGGAATATGAGCCCACTATGGAATATGAGCCCGCTGTGGA 660
DB 636 GAGGCCCAACCAAGGGAATATGAGCCCACTATGGAATATGAGCCCGCTGTGGA 695
QY 661 TATGAGATCCCACTGTGGGGAATATGAGTCCCACTGTGGGGAATATGAGTCCCACTGTGGG 720
DB 696 TATGAGATCCCACTGTGGGGAATATGAGTCCCACTGTGGGGAATATGAGTCCCACTGTGGG 755

QY 721 GGAATATGAGATCCCACTGTGGGGAATATGAGTCCCACTGTGGGGAATATGAGTCCCACT 780
DB 756 GGAATATGAGATCCCACTGTGGGGAATATGAGTCCCACTGTGGGGAATATGAGTCCCACT 815
QY 781 GGGGGAATATGAGATCCCACTGTGGGGAATATGAGTCCCACTGTGGGGAATATGAGTCCCACT 840
DB 816 GGGGGAATATGAGATCCCACTGTGGGGAATATGAGTCCCACTGTGGGGAATATGAGTCCCACT 875
QY 841 CCCCCTGCAATTAAGATCTGTGTGGAATTAAGTGTCTTCAAGATCTTAAGA 898
DB 876 CCCCCTGCAATTAAGATCTGTGTGGAATTAAGTGTCTTCAAGATCTTAAGA 933

RESULT 2
BC022546
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL

BC022546 2267 bp mRNA linear PRI 04-FEB-2002
Homo sapiens, similar to RIKEN cDNA 4930521123 gene, clone
MGC:26816 IMAGE:481804, mRNA, complete cds.
BC022546
BC022546.1 GI:18490706
MGC.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 2267)
Strausberg, R.
Direct Submissions
Submitted (01-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Offices, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mdedpaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LML at: <http://image.llnl.gov>
Series: IRAX Plate: 32 Row: J Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein.
Location/Qualifiers
1. 2267
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:26816 IMAGE:481804"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
32. 961
/codon_start=1
/product="similar to RIKEN cDNA 4930521123 gene"
/protein_id="AAH22546.1"
/db_xref="GI:18490707"
/translation="MAVNSHTSRKALIPSGESVILKRSPPVSELSFPORSBSNVPS
GRKGTFLTSYVVPVSHLVNDPMLSPFMPQGLSDCTIROPFAPNPYIKGTIOAA
PGGMEGQAVFKLSFRKGAIPRAQIAAAGAPIGASVNYPDLSGLYITVP
GAIVSGQFCPAYPIYIGPPPGYVVGKXGKIPBGGAAPGGYGAAPPMYGAAP
VGVGVPBGYGVPPGGYGVPPGGYGAAPGGYGVPPGGYGAAPPGYGAAPPA
NEALPAYEAPSGNTASHSRNTAQDSTLPTTSSS"

PAGSARPHBSTAQA PENBASLPASASSQVHS"
 BASE COUNT 709 a 477 c 469 g 612 t
 ORIGIN

Query Match 58.4%; Score 524.4; DB 9; Length 2267;
 Best Local Similarity 76.4%; Pred. No. 1.2e-120;
 Matches 674; Conservative 0; Mismatches 196; Indels 12; Gaps 2;

```

QY 1 ATGGAGTGAACAGAGCCACCGAGCCGCTGCGGAGCCCTCATCCCTCGGCGAA 60
DB 32 ATGGCGGTGAATCAGAGCCACCGAGAACCGCGGAGCCCTCATCCCTTAACGTGAA 91
QY 61 AGTGTCTTGAACAGTGTGAGATGTGACCTCTGCTCTTCAAGAAACAGTGAATCC 120
DB 92 AGTCTCTTGAAGCGGTCTCCGAATGTGAGGCTCTCTTCCACAGAGATCAGAAAGCTCA 151
QY 121 TATCTCTTAAATGACCAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 152 AATGCTTTAAGTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 211
QY 181 GTGACTTCACTTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 212 ATTAAGTCAATGCTTCAATGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 271
QY 241 AGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 272 AGGAGAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 331
QY 301 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 332 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 391
QY 361 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 392 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 451
QY 421 CCACTTGAAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 452 CCACTTGAAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 511
QY 481 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 512 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 559
QY 541 CCACTTGAAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 570 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 619
QY 601 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 620 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 679
QY 661 TATGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 680 TATGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 739
QY 721 GATATGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 740 GATATGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 799
QY 781 GAGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
DB 800 CTGAGTGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 859
QY 841 CCGGCTGATGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 882
DB 860 CCGGCTGATGAAGTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901

```

RESULT 3
 BC022549 2266 bp mRNA linear PRI 04-FEB-2002
 LOCUS Homo sapiens, similar to RIKEN cDNA 4930521I23 gene, clone
 DEFINITION

ACCSSION MGC:26828 IMAGE:4815849, mRNA, complete cds.

VERSION BC022549
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Strausberg R.
 TITLES Direct Submission
 JOURNAL Submitted (01-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
 COMMENT NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Tohyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdpaxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>
 Series: IRAX Plate: 32 Row: n Column: 3
 This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.
 Location/Qualifiers

FEATURES
 source 1..2266
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:26828 IMAGE:4815849"
 /tissue_type="Brain, hippocampus"
 /clone_id="NIH_MGC_95"
 /lab_host="DH10B"
 /note="Vector: pBluescript"
 32..634
 /codon_start=1
 /product="Similar to RIKEN cDNA 4930521I23 gene"
 /protein_id="AAH22549.1"
 /db_xref="GI:18490711"
 /translation="MAVVGSHTERRRGALLINGESLLKRSFNVSLSPFORSBGSNVFS GRKTGTLPLTSYRVIPLTSCSIDPMLSFMPFDLMTNLTVQOPVPAFNPVKTGLQAA PYGMBGQATKLVFRNGAIEFQQLWAKASAAARGPLTLDMWFSSMGIVYITGK GICALHRLVLQVLLSMBPILQDMBPHLPPTBPHLDMEBNP"

CDS
 BASE COUNT 709 a 477 c 469 g 612 t
 ORIGIN

Query Match 57.8%; Score 519.2; DB 9; Length 2267;
 Best Local Similarity 76.1%; Pred. No. 2.4e-119;
 Matches 671; Conservative 0; Mismatches 196; Indels 13; Gaps 2;

```

QY 1 ATGGAGTGAACAGAGCCACCGAGAGCGGTGCGGAGCCCTCATCCCTCGGCGAA 60
DB 32 ATGGCGGTGAATCAGAGCCACCGAGAACCGCGGAGCCCTCATCCCTTAACGTGAA 91
QY 61 AGTGTCTTGAACAGTGTGAGATGTGACCTCTGCTCTTCAAGAAACAGTGAATCC 120
DB 92 AGTCTCTTGAAGCGGTCTCCGAATGTGAGGCTCTCTTCCACAGAGATCAGAAAGCTCA 151
QY 121 TATCTCTTAAATGACCAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 152 AATGCTTTAAGTGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 211

```


REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

1 (bases 1 to 179222)

Tracey A.

Direct Submission

Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On May 26, 2003 this sequence version replaced gi:31043704.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

----- Project Information

Center project name: bti21021

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 177502 bases at least Q40

Consensus quality: 177971 bases at least Q30

Consensus quality: 178283 bases at least Q20

Insert size: 178722; sum-of-contigs

Insert size: 167315; 12.1% error; agarose-fp

Quality coverage: 7.82x in Q20 bases; sum-of-contigs quality

coverage: 8.59x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 5850: contig of 5850 bp in length

5851 5950: gap of 100 bp

5951 74264: contig of 68314 bp in length

74265 74364: gap of 100 bp

74365 84656: contig of 10292 bp in length

84657 84756: gap of 100 bp

84757 108493: contig of 23737 bp in length

108494 108593: gap of 100 bp

108594 114454: contig of 5861 bp in length

114455 114554: gap of 100 bp

114555 179222: contig of 64668 bp in length.

FEATURES
Source

1. 179222

/organism="Sus scrofa"

/mol_type="Genomic DNA"

/db_xref="taxon:9823"

/clone="PigB-121D21"

/clone_11b="PigB"

1. 5850

/note="assembly fragment:02838

fragment chain:1"

5951. 74264

/note="assembly fragment:01479

fragment chain:1"

74365. 84656

/note="assembly fragment:01113

fragment chain:1"

84757. 108493

/note="assembly fragment:00846

fragment chain:1"

108594. 114454

/note="assembly fragment:01058

fragment chain:1"

114555. 179222

/note="assembly fragment:02416

fragment chain:2"

BASE COUNT 44892 a 44880 c 44117 g 44831 t 502 others

Query Match

26.4% Score 237, DB 2, Length 179222;

Best Local Similarity 73.4% Pred. No. 1.3e-48;
 Matches 303; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 485 CTGCAGTGTCTCTCCACAGACACCTTTCAGCATATTCGATTTATTCATGACCC 544

DB 55353 CTCCAGATGAGAGCCCTCCACCTCCAGATTCAGAGCCCTCCAGATTCAGAGCC 55412

QY 545 CACCAACGATATTCAGTTCACACGAGGATTCAGATTCAGATTCAGATTCAGATTCAG 604

DB 55413 CACCTCCAGATTCAGAGCCCTCCACCTCCAGATTCAGATTCAGATTCAGATTCAG 55472

QY 605 CCGAACGAGGAGATTCAGAGCCCTCCACCTTCAGATTCAGAGCCCTTCAGATTCAG 664

DB 55473 CCGCATTCGCGGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAG 55532

QY 665 GAGTCCCATCTGGGAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAG 724

DB 55533 GAGCTCCATCTGAAAGACACAGACCCCTCCAGATTCAGATTCAGATTCAGATTCAG 55592

QY 725 ATGAGAGCCCATCTGGGAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAG 784

DB 55593 ATGAGAGCCCATCTGCAAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAG 55652

QY 785 GATATGAGCCCATCTGCAAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAG 844

DB 55653 GATATGAGCCCATCTGCAAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAG 55712

QY 845 CTGCATATGAGCTTCATCTGCAAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAG 897

DB 55713 TAGCATATGAGCCCATCTGCAAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAG 55765

RESULT 6

BX470149

LOCUS

DEFINITION Danio rerio clone DKEX-242K7, *** SEQUENCING IN PROGRESS ***; 50

unordered pieces.

BX470149

VERSION BX470149.3 GI:30424228

KEYWORDS HTG PHASRI.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

Burayotai; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission

Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On May 7, 2003 this sequence version replaced gi:30387077.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: zfish-help@sanger.ac.uk

----- Project Information

Center project name: XK242K7

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 197958 bases at least Q40

Consensus quality: 206715 bases at least Q30

Consensus quality: 212351 bases at least Q20

Insert size: 217569; sum-of-contigs

Insert size: 165837; 6.2% error; agarose-fp

Quality coverage: 2.48x in Q20 bases; sum-of-contigs quality

coverage: 3.95x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 50 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 6656: contig of 6656 bp in length
6657 gap of 100 bp
6757 10255: contig of 3499 bp in length
10256 10355: gap of 100 bp
10356 13223: contig of 2868 bp in length
13224 13323: gap of 100 bp
13324 19935: contig of 6612 bp in length
19936 20035: gap of 100 bp
20036 22597: contig of 2562 bp in length
22598 22697: gap of 100 bp
22698 28569: contig of 5872 bp in length
28570 28669: gap of 100 bp
37965 37965: contig of 9296 bp in length
38065 gap of 100 bp
38066 41140: contig of 3075 bp in length
41141 41240: gap of 100 bp
41241 45108: contig of 3868 bp in length
45109 45208: gap of 100 bp
45209 47730: contig of 2522 bp in length
47731 47830: gap of 100 bp
47831 51884: contig of 4054 bp in length
51885 51984: gap of 100 bp
51985 55857: contig of 3873 bp in length
55858 55957: gap of 100 bp
55959 60498: contig of 4541 bp in length
60499 60598: gap of 100 bp
60599 63641: contig of 3043 bp in length
63642 63741: gap of 100 bp
63742 72674: contig of 8933 bp in length
72675 72774: gap of 100 bp
72775 75602: contig of 2828 bp in length
75603 75702: gap of 100 bp
75703 78062: contig of 2360 bp in length
78063 78162: gap of 100 bp
78163 84485: contig of 6323 bp in length
84486 84585: gap of 100 bp
84586 87553: contig of 2968 bp in length
87554 87653: gap of 100 bp
87654 92487: contig of 4834 bp in length
92488 92587: gap of 100 bp
92589 96104: contig of 3517 bp in length
96105 96204: gap of 100 bp
96205 101479: contig of 5275 bp in length
101480 101579: gap of 100 bp
101580 104979: contig of 3400 bp in length
104980 105079: gap of 100 bp
105080 108003: contig of 2924 bp in length
108004 108103: gap of 100 bp
108104 113151: contig of 5048 bp in length
113152 113251: gap of 100 bp
113252 116585: contig of 3334 bp in length
116586 116685: gap of 100 bp
116687 119346: contig of 2661 bp in length
119347 119446: gap of 100 bp
119447 123125: contig of 3679 bp in length
123126 123225: gap of 100 bp
123226 131266: contig of 8041 bp in length
131267 131366: gap of 100 bp
131367 136193: contig of 4827 bp in length
136194 136293: gap of 100 bp
136294 139482: contig of 3189 bp in length
139483 139582: gap of 100 bp
139583 141875: contig of 2293 bp in length
141876 141975: gap of 100 bp
141976 144122: contig of 2147 bp in length
144123 144222: gap of 100 bp
144223 146968: contig of 2746 bp in length
146969 147068: gap of 100 bp

147069 149399: contig of 2331 bp in length
149400 149499: gap of 100 bp
149500 153963: contig of 4464 bp in length
153964 153964: gap of 100 bp
154063 154063: gap of 100 bp
154064 156579: contig of 2516 bp in length
156580 156679: gap of 100 bp
156680 158928: contig of 2249 bp in length
158929 159028: gap of 100 bp
159029 164542: contig of 5514 bp in length
164543 164642: gap of 100 bp
164643 172054: contig of 7412 bp in length
172055 172154: gap of 100 bp
172155 175210: contig of 3056 bp in length
175211 175310: gap of 100 bp
175311 179379: contig of 4069 bp in length
179380 179479: gap of 100 bp
179480 182698: contig of 3219 bp in length
182699 182798: gap of 100 bp
182799 186779: contig of 3981 bp in length
186780 186879: gap of 100 bp
186880 195626: contig of 8747 bp in length
195627 195726: gap of 100 bp
195727 199396: contig of 3670 bp in length
199397 199496: gap of 100 bp
199497 206412: contig of 6916 bp in length
206413 206512: gap of 100 bp
206513 208619: contig of 3107 bp in length
208620 209719: gap of 100 bp
209720 219083: contig of 9364 bp in length
219084 219183: gap of 100 bp
219184 222469: contig of 3286 bp in length.
222469 Location/Qualifiers
1..222469
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone_lib="Dnakey"
1..6656
/note="assembly_fragment:01359
fragment_chain:1"
6757..10255
/note="assembly_fragment:01385
fragment_chain:1"
10356..13223
/note="assembly_fragment:00592
fragment_chain:1"
13324..19935
/note="assembly_fragment:00237
fragment_chain:1"
20036..22597
/note="assembly_fragment:00949
fragment_chain:1"
22698..28569
/note="assembly_fragment:00615
fragment_chain:1"
28670..37965
/note="assembly_fragment:01306
fragment_chain:2"
38066..41140
/note="assembly_fragment:00487
fragment_chain:2"
41241..45108
/note="assembly_fragment:00232
fragment_chain:2"
45209..47730
/note="assembly_fragment:00935
fragment_chain:2"
47831..51884
/note="assembly_fragment:01415
fragment_chain:2"
51985..55857
/note="assembly_fragment:00878

Query Match	Best Local Similarity	Score	DB 2	Length
Matches 303; Conservative	73.4%; Pred. No. 1.3e-48;	26.4%;	DB 2;	222469;
	0; Mismatches 110; Indels 0; Gaps 0;			
misc_feature				
55958..60498				
/note="assembly_fragment:01208				
fragment_chain:3"				
60599..63641				
/note="assembly_fragment:01266				
fragment_chain:3"				
63742..72674				
/note="assembly_fragment:01329				
fragment_chain:3"				
72775..75602				
/note="assembly_fragment:01183				
fragment_chain:4"				
75703..78052				
/note="assembly_fragment:01179				
misc_feature				
485 CTGCAGTGTGCTCTCTCAACAACAACCTTGTCCAGCATATATCAATTTGATCTATGACCCC				
197009 CTCCAGATATACGAGACCCCACTCCAGATATAGGAGCCCACTCCAGATATAGGAGCC				
545 CACCAACGAGATATACAGTCCAAACGAGGGGAAATATGAACTCAACAGAAAGATATAGAG				
197069 CACCTCCAGAGATACGAGAGCCCACTGAAAGACTCAAGACCCCAACAGAGATATAGAG				
605 CCAACAGAGGGGATATAGAGCCCACTATGAGGATATAGAGCCCGCTGTGGGATATAG				
197129 CCCCCTCTGCGGAGATATAGAGCTTCACTCTGCGAGATATAGAGCTTCTCCAGAGATATAG				
665 GAGTCCACCTGAGGGGATATAGAGTCCCACTGAGGGGATATAGAGTCCCACTGAGGGGAT				
197189 GAGCTCACTGAAAGACAGAAAGCCCAACAGAGATATAGAGTCCCACTGAGGGGAT				
725 ATGAGAGCCCCACCTGAGGGGATATAGAGTCCCACTGAGGGGATATAGTGTGCCCACTGGGG				
197249 ATGAAAGCCCCACCTCAAGATATAGAGTCCCGCTTCAAGATATAGAGCTTCACTGAA				
785 GATATGAGAGCCCACTGAGGATATAGAGCCCACTGAGGATATAGAGCCCTTCACTGAG				
197309 GATATGAGAGCCCTCTGAGAGATATAGACACACAGCTGAAATGAAAGCCCACTTC				
845 CTGCATATGAACTCCATCTGCTGAAATACAGCTGCTTCAACAGATCTATG				
197369 TAGCATATGAAAGCCCATCTGCTGAGAAATAGAGCTGCTTCAACATCTG				
RESULT 7				
BX296540/c	129624 bp	DNA	linear	HTG 02-APR-2003
LOCUS				
DEFINITION	Sus scrofa clone PigB-231K18, *** SEQUENCING IN PROGRESS ***, 37			
ACCSSION	BX296540			
VERSION	BX296540.2 GI:29500961			
WORDS	HTG; HTGS PHASRI.			
SOURCE	Sus scrofa (pig)			
ORGANISM	Sus scrofa			
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
	Submitted (02-APR-2003) Wellcome Trust Sanger Institute, Hinxton,			
	Cambridgeshire, CB10 1SA, UK. E-mail enquiries:			
	humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk			
	on Apr 2, 2003 this sequence version replaced gi:29335441.			
	----- Genome Center			
	Center: Wellcome Trust Sanger Institute			
	Center code: SC			
	Web site: http://www.sanger.ac.uk			
	Contact: humquery@sanger.ac.uk			

Project Information	
Center project name: brf33k18	
Summary Statistics	
Assembly: Program: XGAP4; version: 4.5	
Chemistry: Dye-terminator; 100% of reads	
Consensus quality: 113740 bases at least Q40	
Consensus quality: 118187 bases at least Q30	
Consensus quality: 120712 bases at least Q20	
Insert size: 126024; sum-of-contigs	
Insert size: 154528; 4.6% error; agarose-fp	
Quality coverage: 2.35x in Q20 bases; sum-of-contigs	
Quality coverage: 3.19x in Q20 bases; agarose-fp	

NOTE: This is a 'working draft' sequence. It currently	
consists of 37 contigs. The true order of the pieces	
is not known and their order in this sequence record is	
arbitrary. Gaps between the contigs are represented as	
runs of N, but the exact sizes of the gaps are unknown.	
This record will be updated with the finished sequence	
as soon as it is available and the accession number will	
be preserved.	
1	4262: contig of 4262 bp in length
4263	4362: gap of 100 bp
4363	6591: contig of 2629 bp in length
6592	7091: gap of 100 bp
7092	9480: contig of 2389 bp in length
9481	9580: gap of 100 bp
9581	11688: contig of 2108 bp in length
11689	11788: gap of 100 bp
11789	14658: contig of 2870 bp in length
14659	14758: gap of 100 bp
14759	18440: contig of 3682 bp in length
18441	18540: gap of 100 bp
18541	20818: contig of 2278 bp in length
20819	20918: gap of 100 bp
20919	22869: contig of 2051 bp in length
22870	23069: gap of 100 bp
23070	27584: contig of 4515 bp in length
27585	27684: gap of 100 bp
27685	33311: contig of 5527 bp in length
33312	33311: gap of 100 bp
33312	35521: contig of 2210 bp in length
35522	35521: gap of 100 bp
35522	35828: contig of 2307 bp in length
35829	38628: gap of 100 bp
38629	42063: contig of 3435 bp in length
42064	42164: gap of 100 bp
42164	44164: contig of 2001 bp in length
44165	44264: gap of 100 bp
44265	47707: contig of 3443 bp in length
47708	47807: gap of 100 bp
47808	50737: contig of 2930 bp in length
50738	50837: gap of 100 bp
50838	57066: contig of 6229 bp in length
57067	57166: gap of 100 bp
57167	59978: contig of 2812 bp in length
59979	60078: gap of 100 bp
60079	62560: contig of 2482 bp in length
62561	62660: gap of 100 bp
62661	64663: contig of 2203 bp in length
64664	65063: gap of 100 bp
65064	67191: contig of 2128 bp in length
67192	67291: gap of 100 bp
67292	72005: contig of 4714 bp in length
72006	72105: gap of 100 bp
72106	76595: contig of 4490 bp in length
76596	76695: gap of 100 bp
76696	79825: contig of 3130 bp in length
79826	79925: gap of 100 bp
79926	83340: contig of 3615 bp in length
83341	83540: gap of 100 bp
83541	86376: contig of 2636 bp in length
86377	86376: gap of 100 bp

```

* 86377 94620: contig of 8244 bp in length
* 94621 94720: gap of 100 bp
* 94721 97520: contig of 2800 bp in length
* 97521 97620: gap of 100 bp
* 97621 99784: contig of 2164 bp in length
* 99785 99884: gap of 100 bp
* 99885 102112: contig of 2228 bp in length
* 102113 102212: gap of 100 bp
* 102213 104242: contig of 2030 bp in length
* 104243 104342: gap of 100 bp
* 104343 110710: contig of 6368 bp in length
* 110711 110810: gap of 100 bp
* 110811 113367: contig of 2557 bp in length
* 113368 113467: gap of 100 bp
* 113468 115505: contig of 2038 bp in length
* 115506 115605: gap of 100 bp
* 115606 121874: contig of 6269 bp in length
* 121875 121974: gap of 100 bp
* 121975 127096: contig of 5122 bp in length
* 127097 127196: gap of 100 bp
* 127197 129624: contig of 2428 bp in length.

FEATURES
Source
Location/Qualifiers
1..129624
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/clone_lib="PIgR"
/clone_id="PIgR"
1..4262
/feature="assembly_fragment:00361"
fragment_chain:1"
4363..6591
/feature="assembly_fragment:00108"
fragment_chain:1"
7092..9460
/feature="assembly_fragment:00460"
fragment_chain:1"
9581..11688
/feature="assembly_fragment:00959"
fragment_chain:1"
11789..14658
/feature="assembly_fragment:00995"
fragment_chain:2"
14759..18440
/feature="assembly_fragment:00608"
fragment_chain:2"
18541..20818
/feature="assembly_fragment:01101"
fragment_chain:2"
20919..22969
/feature="assembly_fragment:00996"
fragment_chain:2"
23070..27584
/feature="assembly_fragment:00217"
fragment_chain:3"
27685..33211
/feature="assembly_fragment:00019"
fragment_chain:3"
33312..35521
/feature="assembly_fragment:00548"
fragment_chain:3"
35622..38528
/feature="assembly_fragment:00815"
fragment_chain:4"
38629..42063
/feature="assembly_fragment:00246"
fragment_chain:4"
42164..44164
/feature="assembly_fragment:00859"
fragment_chain:5"
44265..47707
/feature="assembly_fragment:00964"
fragment_chain:5"

```

```

misc_feature 47808..50737
/feature="assembly_fragment:01192"
fragment_chain:6"
misc_feature 50838..57066
/feature="assembly_fragment:00542"
fragment_chain:6"
misc_feature 57167..59978
/feature="assembly_fragment:01304"
fragment_chain:7"
misc_feature 60079..62560
/feature="assembly_fragment:00948"
fragment_chain:7"
misc_feature 62661..64963
/feature="assembly_fragment:00016"
/feature="assembly_fragment:00065"
65064..67191
/feature="assembly_fragment:00070"
67292..72005
/feature="assembly_fragment:00070"
72106..76595
/feature="assembly_fragment:00136"
76696..79825
/feature="assembly_fragment:00162"
79926..83540
/feature="assembly_fragment:00367"
83641..86276
/feature="assembly_fragment:00492"
86377..94620
/feature="assembly_fragment:00545"
94721..97520
/feature="assembly_fragment:00745"
97621..99784
/feature="assembly_fragment:00801"

```

```

Query Match 24.9%; Score 223.2; DB 2; Length 129624;
Best Local Similarity 74.2%; Pred. No. 3.9e-45;
Matches 282; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 515 CAGCATATCCATTTGATCTATGACCCCAACCAAGATATGAGTCCACAGAGG 574
DB 65511 CACATAGCCACAGTGTGTATGACCTTACCAATGATGAGGCCAACAGAG 65452

QY 575 AATATGAACTCCACAGAAAGATATGAGCCCAACAGAGGATATGAGCCCACTA 634
DB 65451 AATATGAGCCCACTCCAGATATGAGCCCACTCCAGATATGAGCCCACTC 65392

QY 635 TGGGATATGAGCCCGCCCTGTGATATGAGATCCCACTGGGGATATGAGTCCAC 634
DB 65391 CAGGATATGAGCCCACTCCAGATATGAGCCCACTCCAGATATGAGCCCACT 65332

QY 695 CTGGGGGATATGAGTCCCACTGGGGATATGAGCCCACTGGGGATATGAGTCC 754
DB 65331 CTGAAAGACTCGAAGCCCACTCCAGATATGAGCCCACTGCGGATATGAGCTT 65272

QY 755 CACCTGGGATATGAGTCCCACTGGGGATATGAGCCCACTGCGAGATATGAG 814
DB 65271 CACCTGGGATATGAGTCCCTCCAGATATGAGCTCACTGAAAGACAGAG 65212

QY 815 CCCACCACTGGAATGAAGCCCTTACCTGATATGAAGCTCCATCTGCTGAATA 874
DB 65211 CCCACCACTGGAATGAGTCCCTTCCAGATATGAGCCCACTGCAAGATATG 65152

QY 875 CAGCTGCTCTCAGATCT 894
DB 65151 ATGCCCGCTTCCAGATAT 65132

```

```

RESULT 8
HS250D10 220895 bp DNA linear PRI 05-JUN-2003
DEFINITION Human DNA sequence from clone CTA-250D10 on chromosome 22 contains
the genes for SREBF2 (sterol regulatory element binding
transcription factor 2), NACA (alpha-N-acetylglucosaminidase), a
gene similar to neuronal-specific septin 3, a pseudogene similar to

```

AMT2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a genomic marker D22S1178, a CA repeat polymorphism, ESTs and a Cpg island, complete sequence.
 accession 299716 GI:4456457
 version HTG; AMT2; Cpg Island; D22S1178; NAGA; septin 3; SREBF2.
 keywords Homo sapiens (human)
 source Homo sapiens
 organism
 reference Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 authors 1 (bases 1 to 220895)
 title Clark, G.
 journal Direct Submission
 comment Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Mar 21, 1999 this sequence version replaced gi:4164339.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
 Bm: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C.elegans/wormpep CTA-250D10 is from the human BAC library described in U-J. Kim et al. (1996) Genomics 34, 213-218.
 VECTOR: pBAC108L
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
 This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr22
 This sequence is the entire insert of clone CTA-250D10. The true left end of clone RPI-18601 is at 129979 in this sequence. The true left end of clone RPI-359J16 is at 1335 in this sequence. The true right end of clone RPI-821D11 is at 23458 in this sequence. The true right end of clone RPI-359J16 is at 118711 in this sequence.
 FEATURES
 SOURCE
 1..220895
 /organism="Homo sapiens"
 /mol_type="Genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="22"
 /clone="CTA-250D10"
 /clone_11b="CT978SK-A1"
 17..29
 /note="2.6 copies 5 mer ATTTG 26% conserved"
 repeat_region
 33..325
 /note="AluX repeat: matches 1..301 of consensus"
 repeat_region
 326..338
 /note="2.6 copies 5 mer TTTGA 26% conserved"
 repeat_region
 667..683
 /note="2.1 copies 8 mer AGAGACC 34% conserved"
 repeat_region
 740..751
 /note="2.4 copies 5 mer TTTTA 24% conserved"
 repeat_region
 1443..1457

/note="7.5 copies 2 mer AC 21% conserved"
 repeat_region
 1715..1997
 /note="AluJ repeat: matches 5..293 of consensus"
 repeat_region
 2103..2378
 /note="AluS repeat: matches 2..296 of consensus"
 repeat_region
 2844..2855
 /note="12.0 copies 1 mer A 24% conserved"
 repeat_region
 /note="MIR repeat: matches 70..235 of consensus"
 repeat_region
 complement(3040..3344)
 /note="Alu repeat: matches 1..310 of consensus"
 repeat_region
 3446..3681
 /note="AluM4a repeat: matches 5256..5471 of consensus"
 repeat_region
 3746..3837
 /note="AluM4a repeat: matches 5615..5707 of consensus"
 repeat_region
 3838..4142
 /note="Alu repeat: matches 2..305 of consensus"
 repeat_region
 4143..4441
 /note="AluM4a repeat: matches 5707..5999 of consensus"
 repeat_region
 4442..4747
 /note="AluS repeat: matches 1..292 of consensus"
 repeat_region
 4748..4858
 /note="AluM4a repeat: matches 5999..6110 of consensus"
 repeat_region
 4881..5172
 /note="AluX repeat: matches 1..296 of consensus"
 repeat_region
 5418..5429
 /note="2.4 copies 5 mer GCTGA 24% conserved"
 repeat_region
 complement(5769..6021)
 /note="match: GGS; Bm:A0069839"
 repeat_region
 6074..6089
 /note="2.0 copies 8 mer CCGGCGT 32% conserved"
 repeat_region
 6139..6153
 /note="2.5 copies 6 mer TTTTA 21% conserved"
 repeat_region
 6141..6155
 /note="3.0 copies 5 mer TTTAA 21% conserved"
 repeat_region
 complement(6366..6572)
 /note="MIR repeat: matches 5..214 of consensus"
 repeat_region
 6839..6852
 /note="2.3 copies 6 mer GAAAT 28% conserved"
 repeat_region
 7088..7114
 /note="6.8 copies 4 mer TTTA 36% conserved"
 repeat_region
 7091..7109
 /note="1.9 copies 10 mer ATTATTAC 38% conserved"
 repeat_region
 complement(7122..7418)
 /note="AluX repeat: matches 1..300 of consensus"
 repeat_region
 7464..7502
 /note="19.5 copies 2 mer AT 78% conserved"
 repeat_region
 7502..7524
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region
 7523..7538
 /note="2.0 copies 8 mer CAAATAT 32% conserved"
 repeat_region
 7535..7563
 /note="14.5 copies 2 mer AT 49% conserved"
 repeat_region
 complement(7564..7859)
 /note="Alu repeat: matches 1..296 of consensus"
 repeat_region
 7873..7893
 /note="5.2 copies 4 mer TGT 33% conserved"
 repeat_region
 7914..8200
 /note="AluB repeat: matches 1..290 of consensus"
 repeat_region
 complement(8202..8318)
 /note="AluB/FLM repeat: matches 1..117 of consensus"
 repeat_region
 8387..8405
 /note="11.5 copies 2 mer AC 46% conserved"
 repeat_region
 8483..8498
 /note="2.0 copies 8 mer GTGATGT 32% conserved"
 repeat_region
 8486..8496
 /note="2.2 copies 5 mer GTGT 22% conserved"
 repeat_region
 8606..8622
 /note="2.4 copies 7 mer AGAGAA 34% conserved"
 repeat_region
 complement(8767..8989)
 /note="MIR15 repeat: matches 33..266 of consensus"
 repeat_region
 9004..9014
 /note="2.2 copies 5 mer CTGG 22% conserved"

repeat_region	9037..9140	repeat: matches 17..123 of consensus"
repeat_region	9509..9823	/note="AluX repeat: matches 1..312 of consensus"
repeat_region	9872..9883	/note="2..4 copies 5 mer GTCTT 24% conserved"
gene	10071..49606	/gene="SRBP2"
mRNA	10071..49606	/product="hK250D10.1 (sterol regulatory element binding transcription factor 2)"
	match: CDNA: Em:U02031 Em:U12330 Em:U00968 Em:U09103	
	Em:U16995 Em:U22818 Em:U22819 Em:U12329 Em:AB017337	
	match: EST: Em:R36312 Em:AL040138 Em:H92808 Em:AA053866	
	/evidence="not experimental"	
	join(<10071..10520,11851..12032,14129..14275,17038..17259,18577..18691,18783..18964,20469..20661,21182..21363,23956..24232,28082..28251,36357..36525,38060..38177,40292..40401,41889..42021,43570..43738,46210..46395,48103..48214,48680..49606)	
	/gene="SRBP2"	
	/note="match: proteins: Sw:Q12772 Sw:Q60429 Sw:P36956	
	Sw:Q60416 Tr:O15352"	
	/codon_start=3	
	/evidence="not experimental"	
	/product="hK250D10.1 (sterol regulatory element binding transcription factor 2)"	
	/protein_id="CAB41234.1"	
	/db_xref="GI:5578715"	
	/translation="MLQPSNVQVRFPPDLPSBOLCSFPGSGSGSSGSSSSSSNGRSGSAGVDPSPVQVTVLPSFSPASPOAPLQVPSPTSPPTPRLPILOPRPOPOPOLOLOQOVMTTPSTTPORRITQOPLIYNATAPVLOPOVLOITSSQVPTTIOQOVTOVQAVLTVGTANGTCTATAPVQVAAPOVQVVLVTPQKITSLVLTITKTDSPVMAVONPALTALTTPTOTAPLQVTLVSGSTLTTPVPMGQKRVPIQVQVGVKQLEPPKESGRRITNNIIBRKYSSINDIKIIBKIDVMTDAMHKSQVLEKADIKYIKLOVNHKLQEMVNVKLNQKIKLKGIDLSLVNREYDLKIDFNONVLTMPSPADSGSQAFSPYISDSRSPBLDDAIVKSDPSPALACMWTSSRIICVLTTPCLSPNPLTSLONGGANDSDOHSGSGRSGVSPSSGSGMPMMPTLILMLNVGVLYSVFVGLVHGEVIRPRHSSTVTPRRHKODLDLARDPPAAAGNLOTCLAVIGDALPTSRDLACISLWNVIRISLOLRLVRLKVKVPCRATPATBSGFDEBANTASDADALAVRIHLQHLITGRLPAGSACSDVPMALCAVNLAEKAEKIPSTLVEIHLTAAMGLKTRCGKLGFLASVFLSRAQSLCPREHSAVPDSLRMLCHPIQKRF	
Query Match	24.5%; Score 220; DB 9; Length 220895;	
Best Local Similarity	75.3%; Pred. No. 2.4e-44;	
Matches 274; Conservative	0; Mismatches 90; Indels 0; Gaps 0;	
519	ATATCCAAATGTGATCTATGAGACCCACCAACGATATACAGTCCACAGGAGATA	578
169998	ATTCCTCAATTTCTATGAGACCCACCTGCGAGATATGAGACCCACCTCCGATTA	170057
579	TGGAATCCACAGAGATATGAGACCCACAGGGGATATGAGACCCACCTATGG	638
170058	CGAGACCCACCTGAGATATGAGACCCACCTGAGAAATGAGACCCGCTGTGGG	170117
639	ATATGAGACCCGCTGTGGATATGAGTCCACCTGGGGATATGAGTCCCACTGG	698
170118	ATACAGAGCTCACTGTGCAATATGAGACCCACCTTGTGAATACGAGACCCCACTGC	170177
699	GGATATGAGATCCACCTGGGGATATGAGACCCACCTGGGGATATGAGTCCCACTGC	758
170178	AGATATGAGACCCACCTCTAGATATGAGACCCACCTTGTGAATATGAGACCCCACTC	170237
759	TGGGGATATGATGTCCTCGGGATATGAGACCCACCTGCAAGATATGAGACCC	818
170238	TCTCGGATATGAGACCCACCTCTCGGATATGAGACCCACCTGCAAGATATGAGACCC	170297
QY	819	ACCACTGGAATGAGACCCCTACCTGATATGAGTCCATCTGCTGGAATACAG 878
DB	170298	GCCTGGGGATACAGACCTCACTGCTGATACAGAGGACGCTCAGAAATCTACAG 170357
QY	879	TGCC 882
DB	170358	AGCC 170361
RESULT 9	AC113593	168425 bp DNA linear HTG 06-JUN-2002
AC113593	Mus musculus clone RP23-363124, WORKING DRAFT SEQUENCE, 18 ordered	
DEFINITION	Mus musculus	
ACCESSION	AC113593	
VERSION	AC113593.3	GI:212727432
KEYWORDS	HTG, HTGS PHASE2, HTGS DRAFT, HTGS FULLTOP.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Bukacinska; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 168425)	
JOURNAL	Birren, B., Linton, L., Nusbaum, C. and Lander, E.	
REFERENCE	Mus musculus, clone RP23-363124	
AUTHORS	2 (bases 1 to 168425)	
JOURNAL	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boughalter, B., Brown, A., Camarata, J., Chang, J., Campiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardina, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, M., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Lacombe, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, T., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Plette, R., Pollara, V., Raymond, C., Rette, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, S., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talame, J., Testfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, G., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.	
TITLE	Submitted (03-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
JOURNAL	3 (bases 1 to 168425)	
REFERENCE	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boughalter, B., Brown, A., Camarata, J., Campiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardina, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, M., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Lacombe, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, T., Mihova, T., Mieng, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhthang, P., Plette, R., Pollara, V., Raymond, C., Rette, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, S., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talame, J., Testfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, G., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.	
AUTHORS	Direct Submission	

TITLE
JOURNAL
COMMENT

Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced g1:21313869.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L23816
Center clone name: 363_1_24

Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 16227 bases at least Q40
Consensus quality: 16485 bases at least Q40
Consensus quality: 165972 bases at least Q20
Insert size: 162000; agarose-fp
Insert size: 166725; sum-of-ctnigs
Quality coverage: 8.6 in Q20 bases; agarose-fp
Quality coverage: 8.3 in Q20 bases; sum-of-ctnigs

NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1 1412: contig of 1412 bp in length
1413 1512: gap of 100 bp
1513 3366: contig of 1854 bp in length
3367 3466: gap of 100 bp
3467 5542: contig of 2076 bp in length
5543 5643: gap of 100 bp
5643 7949: contig of 2307 bp in length
7950 8050: gap of 100 bp
8050 42610: contig of 34561 bp in length
42611 42710: gap of 100 bp
42710 45717: contig of 3007 bp in length
45718 45817: gap of 100 bp
45818 50106: contig of 4289 bp in length
50107 50206: gap of 100 bp
50207 52945: contig of 2739 bp in length
52946 53045: gap of 100 bp
53046 56269: contig of 3224 bp in length
56270 56369: gap of 100 bp
56370 59897: contig of 3528 bp in length
59898 59997: gap of 100 bp
59998 67207: contig of 7210 bp in length
67208 67307: gap of 100 bp
67308 78743: contig of 11436 bp in length
78744 78843: gap of 100 bp
78844 89991: contig of 11148 bp in length
89992 90091: gap of 100 bp
90092 107918: contig of 17827 bp in length
107919 108018: gap of 100 bp
108019 124874: contig of 16856 bp in length
124875 124974: gap of 100 bp
124975 142093: contig of 17119 bp in length
142094 142193: gap of 100 bp
142194 167111: contig of 24918 bp in length
167112 167211: gap of 100 bp
167212 168425: contig of 1214 bp in length.

FEATURES
source

Location/Qualifiers
1..168425
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_1fb="RP23-363124"
/clone_1fb="RP23-363124"
1..1412
/note="assembly_fragment"
1513..3366
/note="assembly_fragment"
3367..5542
/note="assembly_fragment"
5543..5643
/note="assembly_fragment"
5644..8050
/note="assembly_fragment"
8051..42610
/note="assembly_fragment"
42611..45717
/note="assembly_fragment"
45718..50106
/note="assembly_fragment"
50107..52945
/note="assembly_fragment"
52946..53046
/note="assembly_fragment"
53047..56269
/note="assembly_fragment"
56270..59897
/note="assembly_fragment"
59898..67207
/note="assembly_fragment"
67208..78743
/note="assembly_fragment"
78744..89991
/note="assembly_fragment"
89992..90091
/note="assembly_fragment"
90092..107918
/note="assembly_fragment"
107919..124874
/note="assembly_fragment"
124875..142093
/note="assembly_fragment"
142094..167111
/note="assembly_fragment"
167112..168425
/note="assembly_fragment"
clone_end:17
vector_side:right

BASE COUNT 44391 a 39572 c 38747 g 44015 t 1700 others
ORIGIN

Query Match 17.5%; Score 156.8; DB 2; Length 168425;
Best Local Similarity 66.4%; Pred. No. 1,7e-28;
Matches 243; Conservative 0; Mismatches 117; Indels 6; Gaps 1;
529 GTGATCTATGAGACCCCAACGAGATATATCACTTCAACGAGGAGATATGAACTCCA 588
66144 GTTCCCTATGAGAGCCCACTCGGGGATATGAGGCTCAACCCGTGGATACGAGTCCCA 66203
589 CCAGAGATATGAGAGCCCAACGAGGAGATATGAGAGCCCACTATGATATGAGAGCC 648
66204 TCTCAGAGATATGAGAGCCCACTCGGGGATATGAGAGCCCACTGAGTATGAGCC 66263
649 CCGCTGTGAGATATGAGAGTCCCACTCGGGGATATGAGATGCCACTGAGGAGATATGAG 708
66264 CCATCTCCGATACGATGCTCTGCTCTCTGAGTACGAG-----CTGTGAGATATGAGC 66317
709 GTCCACCTGAGGAGATATGAGAGCCCACTGAGGAGATATGAGATCCCACTGAGGAGATAT 768
66318 TCCCTCTCTCTCTATATGAGTACCTCCCAATGAGGCTATGAGATGTTCAACTCTGAGATAT 66377
769 GATGAGGAGATATGAGAGCCCACTGAGGAGATATGAGAGCCCACTGAGGAGATATGAGAG 828
66378 GAGCTCAGCTGATGAGATATGAGATCCCACTCTGAGATATGAGAGCCCACTATGAGAG 66437
829 AATGAGCCCTACCCCTGATATGAGATGCTCATCTGCTGAGAAATATACAGCTGCTCTCAC 888

Db 66438 TATGAGCCCAACTCCAGATTAGGAACCTATGGATCCGGTCCCACTCC 66497
Qy 889 AGATCT 894
Db 66498 AGATAT 66503

RESULT 10
AC104325 224086 bp DNA linear ROD 30-MAY-2003
LOCUS Mus musculus clone rp23-204m3 map 15 strain C57BL/6J, complete
DEFINITION sequence.
ACCESSION AC104325
VERSION AC104325.28 GI:31193955
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Mus musculus BAC Clone rp23-204m3
JOURNAL Unpublished
REFERENCE
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Direct Submission
SUBMITTED (10-DEC-2001) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Direct Submission
SUBMITTED (19-MAR-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Jiang, H., Song, L. and Roe, B.A.
TITLE Direct Submission
SUBMITTED (30-MAY-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On May 30, 2003 this sequence version replaced gi:29124187.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

FEATURES
source location/Qualifiers
1..224086
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="15"
/clone="rp23-204m3"
/clone_id="RPC1 - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT 55592 a 55124 c 56095 g 57275 t
ORIGIN

Query Match 17.5%; Score 156.8; DB 10; Length 224086;
Best Local Similarity 66.4%; Pred. No. 1.7e-28;
Matches 243; Conservative 0; Mismatches 117; Indels 6; Gaps 1;

Qy 529 GTGATCTATGAGCCCACTCCAGATTATGACATCCAGAGGAATATGAACTCCA 588
Db 162690 GTTCCCTATGAGCCCACTCCAGATTATGAGGCTTACCCCGGATATCGAGTCCCA 162749

Qy 589 CCAGAAGATATGAGACCCCAACCAAGGGGATATATGAGGCCCACTATGGGATATGAGCC 648
Db 162750 TCTGCAAGGATATGAGACCCCACTCCGCGGATATGAGACCCCACTGATGATATGAGCC 162809
Qy 649 CCGCCTGTGGATATGAGATCCCACTGGGGGATATGAGTCCCACTGGGGGATATGAG 708
Db 162810 CCATCTCCCGATATGAGATCTCTGCTCTCTGATATGAGG-----CTGTGAGATATGAGC 162863
Qy 709 GTCCCACTGGGGGATATGAGACCCCACTGGGGGATATGAGTCCCACTGGGGGATAT 768
Db 162864 TCCCTCTCTCTCTATATATATCTATCCCAATGGGCTATGAGATCTCTCTGATAT 162923
Qy 769 GGTGCCCCCACTGGGGGATATGAGACCCCACTGGAGATATGAGACCCCACTGAG 828
Db 162924 GGAACCTCACTGTGATATGAGATCCCACTCTCGATATGAGACCCCACTATGAG 162983
Qy 829 AATGAAGCCCTTACCCCTGATATGAGATCTGCTGGAATATACAGCTGCTTCAC 888
Db 162984 TATGAGCCCAACTCCAGATTAGGAACCTATGGATCCGGTCCCACTCC 163043
Qy 889 AGATCT 894
Db 163044 AGATAT 163049

RESULT 11
AC107527/c 253149 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-92M24, WORKING DRAFT SEQUENCE, 3
DEFINITION unordered pieces.
ACCESSION AC107527
VERSION AC107527.5 GI:30580771
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS 1 (bases 1 to 253149)
Muzny, D., Marle, M., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, B., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhammed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Bueh, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z.,
Cleveland, C., Cockrell, R., Cox, C., Coyler, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Drapper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C. M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgievski, B., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunnarsson, F., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huylk, S., Hume, J., Idlebird, D., Jackson, A.,
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowls, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenz, L., Louie, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A.,
Mangum, B., Mapua, P., McNeill, T. Z., Meenen, B.,
Mawhinney, S., McLeod, M. P., McNelly, T. Z., Meenen, B.,
Milloreavlevic, A., Miner, G., Minja, B., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muntaner, M., Murphy, M., Nair, L.,
Nankervill, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwankwelen, O., Okunolu, G., Olarnpungoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfamkoeh, C.,

Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puato, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodery, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Sma, J., Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Soza, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trajce, Z., Umami, K., Valdes, R., Vera, V., Villaseca, D., Waldron, J., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 253149)
Worley, K.C.

Direct Submission
Submitted (23-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 253149)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:23664681.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GKFL
Center clone name: CH230-92M24

----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 242817 bases at least Q40
Consensus quality: 245517 bases at least Q20
Consensus quality: 247516 bases at least Q20
Estimated insert size: 257092; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

* NOTE: Baited insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 250781: contig of 250781 bp in length
* 250782 250881: gap of unknown length
* 250882 251966: contig of 1085 bp in length

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature

Query Match
Best Local Similarity 65.9%; Pred. No. 3.1e-27;
Matches 238; Conservative 0; Mismatches 117; Indels 6; Gaps 1;

BASE COUNT 64040 a 61494 c 61630 g 62235 t 3750 others
ORIGIN

Query 534 CTATGACCCCAACCAAGATATGACGTCACCAAGGGGATATGACATCCACGAA 593
Db 118121 CTATGACCCCAACCAAGATATGACGTCACCAAGGGGATATGACATCCACGAA 118062

Query 594 AGATATGACCCCAACCAAGGAGATATGACCCCAAGATATGACCCCAAG 653
Db 118061 AGATATGACCCCAACCAAGGAGATATGACCCCAAGATATGACCCCAAG 118002

Query 654 TGTGGATATGACGTCACCAAGGAGATATGACCCCAAGATATGACCCCAAG 713
Db 118001 TGTGGATATGACGTCACCAAGGAGATATGACCCCAAGATATGACCCCAAG 117948

Query 714 ACGTGGGAGATATGACCCCAAGGAGATATGACCCCAAGATATGACCCCAAG 773
Db 117947 TGTCTCTATGACGTCACCAAGGAGATATGACCCCAAGATATGACCCCAAG 117888

Query 774 CCAAGGAGATATGACCCCAAGGAGATATGACCCCAAGATATGACCCCAAG 833
Db 117887 CCAAGGAGATATGACCCCAAGGAGATATGACCCCAAGATATGACCCCAAG 117828

Query 834 AGCCCTACCCCTGATATGACGTCACCAAGGAGATATGACCCCAAGATATGACCCCAAG 893
Db 117827 GCGCCCACTCTATGACGTCACCAAGGAGATATGACCCCAAGATATGACCCCAAG 117768

Query 894 T 894
Db 117767 T 117767

RESULT 12
AC132969/c 270171 bp DNA linear HTG 20-NOV-2002

LOCUS AC132969
DEFINITION Rattus norvegicus clone CH230-327L20, WORKING DRAFT SEQUENCE.
ACCESSION AC132969
VERSION AC132969.3 GI:25139203
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 270171)
Muzny, D., Marle, M., Metzger, M., Lee, A., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooke, S., Amin, A., Angiano, D., Ayalew, V., Ayagi, A., Ayodeji, M., Baca, B., Baden, D., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benamed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabriel, A., Ganta, R., Garcia, A., Garner, T., Garza, W., George, R., Geier, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huliyil, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovat, C., Kowals, C., Kraft, C. L., Lebow, H., Levan, U., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenz, H., Louisa, L., Louisa, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapus, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muidasa, M., Murphy, M., Nair, L., Narkewicz, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nsamenang, O., Okunju, G., Olamugbo, A., Pal, S., Parks, K., Patel, S., Paul, H., Perez, A., Perez, J., Pfannkuch, C., Plopper, F., Polidexter, A., Popovic, D., Prymus, S., Pu, L., L., Plazo, M., Quinz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Rella, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shateman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Slason, I., Sitter, C. D., Smaje, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Swatek, A., Tabot, P., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Yamlor, K., Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleciyk, R., Woodden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, D., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

TITLE JOURNAL REFERENCE

2 (bases 1 to 270171)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (05-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 270171)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE AUTHORS TITLE JOURNAL COMMENT

On Nov 20, 2002 this sequence version replaced gi:23308490.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contig within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KB0D
Center clone name: CH230-337L20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 251525 bases at least Q40
Consensus quality: 254049 bases at least Q30
Consensus quality: 255599 bases at least Q20
Estimated insert size: 259926; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a "working draft" sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 270171: contig of 270171 bp in length.
Location/Qualifiers
1..270171
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-337L20"
1..2686
/note="wgs end-extension
clone_end:5p6"
4191..5073
/note="clone boundary
clone_end:5p6
site:
end_sequence:BZ160678"
complement(267941..268827)
/note="clone boundary
clone_end:T7
site:
end_sequence:BZ160665"

BASE COUNT 67733 a 63574 c 62822 g 63249 t 12793 others
ORIGIN

Query Match 16.9%; Score 151.8; DB 2; Length 270171;
Best Local Similarity 65.9%; Pred. No. 3.1e-27;
Matches 238; Conservative 0; Mismatches 117; Indels 6; Gaps 1;

534 CTATGACCCCAACAGATATACAGTCCACAGGAGATATGAACTCCACAGA 593
15369 CTATGAGCCCAACAGATATACAGTCCACAGGAGATATGAACTCCACAGA 15310
594 AGATATGAGCCCAACAGATATACAGTCCACAGGAGATATGAACTCCACAGA 653
15309 AGATATGAGCCCAACAGATATACAGTCCACAGGAGATATGAACTCCACAGA 15250
654 TGTGGATATGAGTCCCACTGGAGATATGAGTCCCACTGGAGATATGAGTCC 713
15249 TCCAGATATGAGTCCCTGCTTCTGATACGAG-----CTCCAGATATGATCC 15196
714 ACCTGGAGATATGAGTCCCACTGGAGATATGAGTCCCACTGGAGATATGAGTCC 773
15195 TCCTCTCTATACGATACCTACCTATGAGGATATGAGTCCCACTCTGGATATGAGC 15136

QY 774 CCCACTGGGGATATGAGCCCACTGACGATATGAGAGCCCAAGCTGGAATGA 833
 DB 15135 CCCACTGGAGATAGCATGCCCACTCTGATACGAAGCCCACTATGACATATG 15076
 QY 834 AGCCCTACCCCTGATATGAGCTTCATCTGCGAATATGAGCTGCTTCAAGATC 893
 DB 15075 CCCCCCACTCTACATAGTACCTATGAAATCCGGTTCCCACTCCCGAGATA 15016
 QY 894 T 894
 DB 15015 T 15015

RESULT 13
 AF499026 793 bp mRNA linear ROD 23-APR-2002
 LOCUS Rattus norvegicus WW-domain binding protein 2 (Wbp2) mRNA, complete cds.
 DEFINITION
 ACCESSION AF499026 GI:20269980
 VERSION AF499026.1 GI:20269980
 KEYWORDS
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 793)
 AUTHORS Zaminin, M. and Nitsch, R.
 JOURNAL Isolation of the cDNA encoding for rat Wbp2
 TITLE Unpublished
 REFERENCE 2 (bases 1 to 793)
 AUTHORS Zaminin, M. and Nitsch, R.
 JOURNAL Direct Submission
 TITLE Submitted (04-APR-2002) DBPC, University of Naples Federico II, via Pansini 5, Naples, NA 80131, Italy
 FEATURES
 source
 1. 793
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 1..793
 /gene="Wbp2"
 1..789
 /gene="Wbp2"
 /codon_start=1
 /product="WW-domain binding protein 2"
 /protein_id="AA018132.1"
 /db_xref="GI:20269980"
 /translation="MALNKNHSGGGVIVNTBSTILMSYDHYELTFNDKRVPEAFK
 TKGTIVLTPRVIFLSKGDMSFPMFVIMKCEVQPVGFANPIKGVKAAAG
 GMEGASAYKLTFTAGALIFGQMLQVASQASRGVPMGAYGYPMPSGAYVPPVA
 NGMYPCPGYPP
 AAKAAEAAAYVNPGRNHYVPTSGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
 BASE COUNT 189 a 259 c 196 g 149 t

ORIGIN
 Query Match 14.3%; Score 128.4; DB 10; Length 793;
 Best Local Similarity 60.3%; Pred. No. 2.6e-21;
 Matches 251; Conservative 0; Mismatches 156; Indels 9; Gaps 2;

QY 1 ATGCGATGACCAAGACCAACCGAGCCGCTGCGGGCCCTCATCCCTCTGGCGAA 60
 DB 1 ATGGCGCTTCACAAAGATCATCTCAGAGCGCGCGAGTGCATCTCAACCACT--GAG 57
 QY 61 AGTGTCTGAAGCAGTGAAGATGAGACCTCTGCTCTTACAGAAACAGTGAATGC 120
 DB 58 AGCATCTTAATGCTCTATATCATGTGGAATTAATTAACGACATGAAGATGACCA 117
 QY 121 TATCTCTTAATGACCAAGAAAGAGAGCTGTTCTCACTTACATACCGGGTGTCTTC 180
 DB 118 GAGGCTTCAAGAGGACCAAGAAAGGACCGTCTACCTAATCCATACCGGGTGTCTTT 177
 QY 181 GTGACTTCACTTATGTCAATGACCCCATGCTTTCTTTATGATGCGCTTGCTGATG 240

DB 178 CTGTC-----CAAGGAGAGATGCCATGGGCTCTTCATGATGCCCTTCTACCTGATG 231
 QY 241 ACTGACTGACCATTTGAACAACCAATTTTGGCCCCCACTACATTAAGAAACCATTCAG 300
 DB 232 AAGACTGTGAGGTCAACACACCGCGTGTGGTGCCACTTCATTAAGGGGACACTGAAA 291
 QY 301 GCAGCTCAGAGTGGGCTGGGAGAGACAGCTGTTTAAATTAATCTCTTGAGAAAGA 360
 DB 292 GCTGAAGCAGAGAGGTGGTGGAGAGCTCTGCTCTCTCAAGAGTACCTTCAACAGAGG 351
 QY 361 GGTGCCATGCAATTTGCCCACTGATGTTAAAGCTGCTGCTGCTGCACAGAG 416
 DB 352 GGTCCATTTGAGTTGGGCGACGATGCTCCAGAGTGGATCTCAAGCTCTCCAGAG 407

RESULT 14
 M040826 1802 bp mRNA linear ROD 04-MAR-2003
 LOCUS Mus musculus WW-domain binding protein 2 mRNA, complete cds.
 DEFINITION
 ACCESSION M040826 GI:1777578
 VERSION U40826
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS Sudol, M., Chen, H.I., Bougeret, C., Blinbond, A. and Bork, P.
 JOURNAL Characterization of a novel protein-binding module--the WW domain
 TITLE FEBS Lett. 369 (1), 67-71 (1995)
 MEDLINE 95369475
 PUBMED 7641887
 REFERENCE 2 (bases 1 to 1802)
 AUTHORS Chen, H.I. and Sudol, M.
 JOURNAL The WW domain of Yes-associated protein binds a proline-rich ligand
 TITLE that differs from the consensus established for Src homology
 3-binding modules
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92 (17), 7819-7823 (1995)
 MEDLINE 95372370
 PUBMED 7644498
 REFERENCE 3 (bases 1 to 1802)
 AUTHORS Chen, H.I. and Sudol, M.
 JOURNAL Identification and Characterization of Protein Ligands to the WW
 TITLE Domain by Western Ligand Blotting
 JOURNAL (in) Marshak, D.R. (Ed.),
 TECHNIQUES IN PROTEIN CHEMISTRY VII: 3-12;
 Academic Press, Inc., USA (1996)
 4 (bases 1 to 1802)
 AUTHORS Chen, H.I. and Sudol, M.
 JOURNAL Direct Submission
 TITLE Submitted (17-NOV-1995) Henry I. Chen, Laboratory of Molecular
 Oncology, The Rockefeller University, 1230 York Avenue, New York,
 NY 10021, USA
 FEATURES
 source
 1..1802
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /dev_stage="16 day embryo"
 40..825
 /note="WBP-2; putative protein ligand to the WW domain of
 yes-kinase associated protein YAP"
 /codon_start=1
 /product="WW-domain binding protein 2"
 /protein_id="AAB40893.1"
 /db_xref="GI:1777579"
 /translation="MALNKNHSGGGVIVNTBSTILMSYDHYELTFNDKRVPEAFK
 TKGTIVLTPRVIFLSKGDMSFPMFVIMKCEVQPVGFANPIKGVKAAAG
 GMEGASAYKLTFTAGALIFGQMLQVASQASRGVPMGAYGYPMPSGAYVPPVA
 NGMYPCPGYPP
 AAKAAEAAAYVNPGRNHYVPTSGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
 BASE COUNT 396 a 565 c 437 g 404 t

ORIGIN

Query Match 14.3%; Score 128.4; DB 10; Length 1802;
 Best Local Similarity 60.3%; Pred. No. 2.5e-21;
 Matches 251; Conservative 0; Mismatches 156; Indels 9; Gaps 2;

Qy 1 ATGCGATGAAACAGACCAACACGAGAGCGTGTGGGAGCCCTCATCCCTTGAGGAA 60
 Db ATGCGGCTCAACAGATCACTCAGAGGGGGCGAGATGATTCCTCAACACACT---GAG 96
 Qy 61 AGTGTCTTGAAGCAGTGTGAGATGTGACCTCTGCTTCTTCAAGAAACAGTGAATCC 120
 Db AGCATCTTAATGTCTTATGATCATGTGAGAGCTTACCTTCAACGACAGAAAGATGCCA 156
 Qy 121 TATCTCTTAAATGCAAAAGAAAGAAAGCTTTTCTCACTTCATCCGGGTGCTTC 180
 Db GAGGCTTCAAAAGGACCAAGAAAGGACCGCTTCACTTACTCCGTACCGGGTATCTTT 216
 Qy 181 GTGACTTCACACTTAATGATGACCCCATGCTTCTTTATGATGCGGTTGGCTGATG 240
 Db CTGTCTTAAG-----GGAAAGAGCGCATGCAATGCTTCAATGATGCCCTTTTACTGATG 270
 Qy 241 AGTACTGACACATTAACAACAATTTTGGCCCACTACATTAAGAAACCATTCAG 300
 Db AAGGACTGTGATTAAGACAGCCGCTGTTGTGCGAATTCATTAAGGAAATGTGAAG 330
 Qy 301 GCAGCTCAGAGTGTGCTGGGAAAGCAAGCTTTTAAATATCTTCAAGAAAGGA 360
 Db GCTGAACAGAGAGGTGCTGGAAAGGCTCCGCTTCAACAAGCTTCAACAGCAGGG 390
 Qy 361 GGTGCGATGCAATTTGGCCCACTGATGTAAAGCTGCTGCTGCTGCAGAGG 416
 Db GGGCCCATTAAGTTGGGACAGAGATGCTCAAGGTGCAATCTCAAGCTTCAAGAG 446

RESULT 15
 BX470149/c 222469 bp DNA linear HTG 07-MAY-2003
 LOCUS BX470149.3 GI:30424228
 DEFINITION Danio rerio clone DKEX-242K7, *** SEQUENCING IN PROGRESS ***, 50
 unoriented pieces.

ACCESSION BX470149
 VERSION BX470149.3 GI:30424228
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio

Bakeryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 1 (bases 1 to 222469)

REFERENCE
 AUTHORS Burton, J.
 TITLES Direct Submission
 JOURNAL Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On May 7, 2003 this sequence version replaced gi:30387077.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information
 Center project name: zK242K7
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 197958 bases at least Q40
 Consensus quality: 206715 bases at least Q30
 Consensus quality: 212351 bases at least Q20
 Insert size: 217569; sum-of-contigs
 Insert size: 165837; 6.2% error; agarose-fp
 Quality coverage: 2.48x in Q20 bases; sum-of-contigs Quality
 coverage: 3.95x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 50 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 6656: contig of 6656 bp in length
 6657 gap of 100 bp
 6757 10255: contig of 3499 bp in length
 10256 gap of 100 bp
 10355 13223: contig of 2866 bp in length
 13224 gap of 100 bp
 13324 19935: contig of 6612 bp in length
 19936 gap of 100 bp
 20036 22597: contig of 2562 bp in length
 22598 gap of 100 bp
 22599 28570: contig of 5872 bp in length
 28571 gap of 100 bp
 28572 37965: contig of 9296 bp in length
 37966 gap of 100 bp
 38066 41140: contig of 3075 bp in length
 41141 gap of 100 bp
 41241 45108: contig of 3868 bp in length
 45109 gap of 100 bp
 45209 47730: contig of 2522 bp in length
 47731 gap of 100 bp
 47831 51884: contig of 4054 bp in length
 51885 gap of 100 bp
 51985 55857: contig of 3873 bp in length
 55858 gap of 100 bp
 55958 60498: contig of 4541 bp in length
 60499 gap of 100 bp
 60599 63641: contig of 3043 bp in length
 63642 gap of 100 bp
 63742 72674: contig of 8933 bp in length
 72675 gap of 100 bp
 72775 75602: contig of 2826 bp in length
 75603 gap of 100 bp
 75703 78062: contig of 2360 bp in length
 78063 gap of 100 bp
 78163 84485: contig of 6323 bp in length
 84486 gap of 100 bp
 84586 87553: contig of 2966 bp in length
 87554 gap of 100 bp
 87654 92487: contig of 4834 bp in length
 92488 gap of 100 bp
 92588 96104: contig of 3517 bp in length
 96105 gap of 100 bp
 96205 101479: contig of 5275 bp in length
 101480 gap of 100 bp
 101580 104979: contig of 3400 bp in length
 104980 gap of 100 bp
 105080 108003: contig of 2924 bp in length
 108004 gap of 100 bp
 108104 113151: contig of 5046 bp in length
 113152 gap of 100 bp
 113251 116585: contig of 3334 bp in length
 116586 gap of 100 bp
 116587 119346: contig of 2661 bp in length
 119347 gap of 100 bp
 119447 123125: contig of 3679 bp in length
 123126 gap of 100 bp
 123226 131265: contig of 8041 bp in length
 131267 gap of 100 bp
 131367 136193: contig of 4827 bp in length
 136194 gap of 100 bp
 136293 139482: contig of 3189 bp in length
 139483 gap of 100 bp
 139582 141875: contig of 2293 bp in length
 141876 gap of 100 bp
 141975 144122: contig of 2147 bp in length

* 144123 144222: gap of 100 bp
* 144223 14668: contig of 2746 bp in length
* 14669 147068: gap of 100 bp
* 147069 149399: contig of 2331 bp in length
* 149400 149499: gap of 100 bp
* 149500 153963: contig of 4466 bp in length
* 153964 154063: gap of 100 bp
* 154064 156579: contig of 2516 bp in length
* 156580 156679: gap of 100 bp
* 156680 158928: contig of 2249 bp in length
* 158929 159028: gap of 100 bp
* 159029 164542: contig of 5511 bp in length
* 164543 164642: gap of 100 bp
* 164643 172054: contig of 7412 bp in length
* 172055 172154: gap of 100 bp
* 172155 175210: contig of 3056 bp in length
* 175211 175310: gap of 100 bp
* 175311 179379: contig of 4069 bp in length
* 179380 179479: gap of 100 bp
* 179480 182698: contig of 3219 bp in length
* 182699 182798: gap of 100 bp
* 182799 186779: contig of 3981 bp in length
* 186780 186879: gap of 100 bp
* 186880 195626: contig of 8747 bp in length
* 195627 195726: gap of 100 bp
* 195727 199396: contig of 3670 bp in length
* 199397 199496: gap of 100 bp
* 199497 206412: contig of 6916 bp in length
* 206413 206512: gap of 100 bp
* 206513 209619: contig of 3107 bp in length
* 209620 209719: gap of 100 bp
* 209720 219083: contig of 9364 bp in length
* 219084 219183: gap of 100 bp
* 219184 222469: contig of 3286 bp in length.
Location/Qualifiers
1. 222469

source
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone_id="DKX-242K7"
/clone_idb="DanioKey"
1. 6656
/note="assembly_fragment:01359
fragment_chain:1"
misc_feature
6757..10255
/note="assembly_fragment:01385
fragment_chain:1"
misc_feature
10356..13223
/note="assembly_fragment:00592
fragment_chain:1"
misc_feature
13324..19935
/note="assembly_fragment:00237
fragment_chain:1"
misc_feature
20036..32597
/note="assembly_fragment:00949
fragment_chain:1"
misc_feature
22698..28569
/note="assembly_fragment:00615
fragment_chain:1"
misc_feature
28670..37965
/note="assembly_fragment:01306
fragment_chain:2"
misc_feature
38066..41140
/note="assembly_fragment:00487
fragment_chain:2"
misc_feature
41241..45108
/note="assembly_fragment:00232
fragment_chain:2"
misc_feature
45209..47730
/note="assembly_fragment:00935
fragment_chain:2"
misc_feature
47831..51884
/note="assembly_fragment:01415

fragment_chain:2"
misc_feature
51985..55857
/note="assembly_fragment:00878
fragment_chain:3"
misc_feature
55958..60498
/note="assembly_fragment:01208
fragment_chain:3"
misc_feature
60599..63641
/note="assembly_fragment:01266
fragment_chain:3"
misc_feature
63742..72674
/note="assembly_fragment:01329
fragment_chain:3"
misc_feature
72775..75602
/note="assembly_fragment:01183
fragment_chain:4"
misc_feature
75703..78062
/note="assembly_fragment:01179

Query Match 13.7%; Score 132.8; DB 2; Length 222469;
Best Local Similarity 88.7%; Pred.No.5.8e-20;
Matches 133; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 171 GGTGGCTTCGTGACTGCACTTAGTCATGATGACCCCATGCTTTATGATGCCGTT 230
DB 190751 GGTGATCTTCGTGACTGCACTGACGACGACACCCCATGTTTCTTCATGATGCCATT 190692
QY 231 TGCCGTGATGATGACTGACCATTTGAACACCAATTTTGCCCCCACTACATTAAAG 290
DB 190691 TGATCTGATGATGACTGACCATTTGAACCAACCCGCTTTTGCCCCCACTACATTAAAG 190632
QY 291 AACCATTCAGGAGCTCCAGGTGGTGGCTG 320
DB 190631 AACCGTTCAGGAGCTCCAGATGATGATG 190602

Search completed: December 16, 2003, 15:10:10
Job time : 3343.84 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 05:54:30 / Search time 239.646 Seconds

(Without alignments) 10115.338 Million cell updates/sec

Title: US-09-864-291-4_COPY_36_933

Perfect score: 898

Sequence: 1 atgcgcagtgacacagagccca.....tgcctctcacagatctatga 898

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Genseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	898	100.0	1413	24	AAS20601	DNA encoding bovin
2	374.4	41.7	1001	22	AAS20602	DNA encoding human
3	262.6	29.2	467	22	ABL01229	Human reproductive
4	262.6	29.2	467	23	ABL96688	Human testicular a
5	220	24.5	7099	22	ABL04882	Human reproductive
6	220	24.5	7099	23	ABL97776	Human testicular a
7	220	24.5	220895	24	ABR84798	Human cDNA differe
8	216.8	24.1	436	22	ABA67850	Human foetal liver

9	216.8	24.1	436	22	AAK42003	Human bone marrow
10	216.8	24.1	436	22	AAI48070	Probe #16756 used
11	216.8	24.1	436	24	ABR16034	Human genome-deriv
12	195.4	21.8	321	23	AAI01355	Human reproductive
13	195.4	21.8	321	23	ABL96688	Human testicular a
14	122	13.6	1885	24	ABR70430	Human bone remodel
15	122	13.6	1915	24	AAE22363	Human secreted pro
16	122	13.6	1915	25	ABR73614	Secreted protein-e
17	122	13.6	1915	25	ABR15681	Human secreted pro
18	122	13.6	1915	25	ABR67208	Human secreted pro
19	121.2	13.5	894	23	AAS70582	DNA encoding novel
20	114.4	12.7	899	25	ABX41819	Bovine EST associa
21	103.8	11.6	454	21	AAO04014	Human secreted pro
22	92.4	10.3	291	22	AAI04683	Human secreted pro
23	92.4	10.3	291	23	ABR97777	Human reproductive
24	86	9.6	471	22	ABR55246	Human testicular a
25	86	9.6	471	22	AAK28957	Human bone marrow
26	86	9.6	471	22	AAI34906	Probe #1592 used t
27	86	9.6	471	22	ABR03491	Human genome-deriv
28	73.6	8.2	420	24	ABR76335	Bacillus lichenifo
29	70	7.8	136	21	AAK2819	Human secreted pro
30	68.8	7.7	591	22	ABR49404	Human breast cell
31	68.8	7.7	591	22	ABR67315	Human foetal liver
32	68.8	7.7	591	22	ABR34407	Probe #12873 for g
33	68.8	7.7	591	22	AAK15750	Human brain expres
34	68.8	7.7	591	22	AAK41489	Human bone marrow
35	68.8	7.7	591	22	AAI22235	Probe #12168 for g
36	68.8	7.7	591	22	AAI07935	Probe #1926 used t
37	68.8	7.7	591	22	AAI47531	Probe #7926 used t
38	68.8	7.7	591	23	ABR41079	Human liver singl
39	68.8	7.7	591	24	ABR15494	Human genome-deriv
40	68.8	7.7	1959	22	ABR44250	Human breast cell
41	68.8	7.7	1959	22	ABR54639	Human foetal liver
42	68.8	7.7	1959	22	ABR24480	Probe #2946 for ge
43	68.8	7.7	1959	22	AAK02984	Human brain expres
44	68.8	7.7	1959	22	AAK28431	Human bone marrow
45	68.8	7.7	1959	22	AAI12997	Probe #2930 for ge

ALIGNMENTS

RESULT 1	
AAI20601	
ID	AAS20601 standard; cDNA, 1413 BP.
XX	
AC	AAS20601;
XX	
DT	09-APR-2002 (first entry)
XX	
DE	DNA encoding bovine perinuclear theca 32 (PT32).
XX	
KM	Testicular WW domain binding protein; hWWBP; perinuclear theca 32;
KM	PT32; contraceptive; fertility; oocyte activation; vaccine;
KM	globozoospermy; spermiogenesis; spermatozoa; tyrosine kinase; c-Yes;
KW	immunoreceptor; bovine; gene; ss.
OS	Bos sp.
XX	
PH	Key
FT	primer_bind
FT	Location/Qualifiers
FT	30..50
FT	/*tag= a
FT	/note= "Primer binding site for cDNA isolation. The
FT	sequence differs from that of the forward primer
FT	given in AAS20603"
FT	
FT	CDS
FT	36..977
FT	/*tag= b
FT	/product= "PT32"
FT	/note= "Perinuclear theca 32"
FT	978..1001
FT	/*tag= c
FT	/note= "Primer binding site for cDNA isolation. The
FT	sequence differs from that of the reverse primer

FT XX given in AAS20604"

XX FN WO200190185-A2.

XX PD 29-NOV-2001.

XX PE 25-MAY-2001; 2001MO-CA00738.

XX PR 25-MAY-2000; 2000CA-2307128.

XX PR 25-MAY-2000; 2000US-206979P.

XX PA (TOOH) UNIV QUEBENS KINGSTON.

XX PA (UYOR-) UNIV OREGON HEALTH SCI.

XX P1 Oko R, Sutovsky P;

XX XW MPI; 2002-097644/13.

XX DR P-PSDB; AAU74604.

XX PT Isolated perinuclear theca 32 polypeptide that interacts with activated

XX PT tyrosine kinase c-Yes, for enhancing fertility; treating/diagnosing

XX PT diminished fertility and abnormal spermiogenesis and for providing

XX PT contraception -

XX PS Claim 10; Fig 2A-B; 103bp; English.

XX CC The invention describes an isolated perinuclear theca 32 (PT32)

XX CC polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is

XX CC useful for: enhancing fertility in a mammal; treating globozoospermy, by

XX CC expressing (I) in spermatozoa; inhibiting fertilization, by introducing

XX CC (I) or its antigenic fragment into a mammal to elicit an immune

XX CC response; enhancing the ability of round spermatids to activate oocytes;

XX CC treating or diagnosing diminished fertility and abnormal spermiogenesis;

XX CC in providing oocyte activation; identifying contraceptive and

XX CC fertility-enhancing agents. The polynucleotide is useful for producing

XX CC (I) by recombinant techniques, as vaccine, as diagnostic reagents, and

XX CC for chromosome identification. An antibody against (I) is useful in

XX CC immunological assays, in immunoprecipitative methods, to identify cells

XX CC expressing (I), and to purify (I) by affinity chromatography. A

XX CC transgenic animal is useful as an animal model for studying human

XX CC fertility and reproductive biology, and for screening compounds to

XX CC identify modulators of oocyte activation. The use of (I) prevents the

XX CC entry of components which are detrimental to embryonic development into

XX CC the oocyte during oocyte activation with crude sperm extract and avoids

XX CC the propagation of viruses such as HIV (human immunodeficiency virus) and

XX CC SIV (Simian immunodeficiency virus) carried in the sperm. This sequence

XX CC encodes the bovine perinuclear theca 32 (PT32), described in the method

XX CC of the invention.

XX SQ Sequence 1413 BP; 377 A; 363 C; 369 G; 304 T; 0 other;

Query Match 100.0%; Score 898; DB 24; Length 1413;

Best Local Similarity 100.0%; Pred. No. 5,76-257;

Matches 898; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCAATGAAACAGAGCCAGACCGGCTGTGGGCGCTTACCTCTTGGCGAA 60

DB 36 ATGGCAATGAAACAGAGCCAGACCGGCTGTGGGCGCTTACCTCTTGGCGAA 95

QY 61 AGTGTCTTGAAGCAATGAGATGTGAGACCTTCTCTTACAGAAACAGATGGAATCC 120

DB 96 AGTGTCTTGAAGCAATGAGATGTGAGACCTTCTCTTACAGAAACAGATGGAATCC 155

QY 121 TATCTCTTAAATGCAAAAGAAAGAGTGTCTTCACTTCAATCCGGGTGTCTTC 180

DB 156 TATCTCTTAAATGCAAAAGAAAGAGTGTCTTCACTTCAATCCGGGTGTCTTC 215

QY 181 GTGACTTCACTTAAATGCAATGAGCCCAATGCTTCTTAAATGAGCCGTTGGCTGATG 240

DB 216 GTGACTTCACTTAAATGCAATGAGCCCAATGCTTCTTAAATGAGCCGTTGGCTGATG 275

QY 241 AGTGAATGCACTTAAATGCAATGAGCCCAATGCTTCTTAAATGAGCCGTTGGCTGATG 300

DB 276 AGTGAATGCACTTAAATGCAATGAGCCCAATGCTTCTTAAATGAGCCGTTGGCTGATG 335

QY 301 GCACTTCACTTAAATGCAATGAGCCCAATGCTTCTTAAATGAGCCGTTGGCTGATG 360

DB 336 GCACTTCACTTAAATGCAATGAGCCCAATGCTTCTTAAATGAGCCGTTGGCTGATG 395

QY 361 GGTGCAATGCAATGAGCCCAATGCTTCTTAAATGAGCCGTTGGCTGATG 420

DB 396 GGTGCAATGCAATGAGCCCAATGCTTCTTAAATGAGCCGTTGGCTGATG 455

QY 421 CCACTTGAAGTGTAAATTAATGCTTCACTTCAATGAGCCGTTGGCTGATG 480

DB 456 CCACTTGAAGTGTAAATTAATGCTTCACTTCAATGAGCCGTTGGCTGATG 515

QY 481 GGGGCTGAGTGTCTTCACTTCAATGAGCCGTTGGCTGATG 540

DB 516 GGGGCTGAGTGTCTTCACTTCAATGAGCCGTTGGCTGATG 575

QY 541 CCCCCACCAAGATATATAGTCCACAGAGGGGATATAGAGCTCCACAGAGATAT 600

DB 576 CCCCCACCAAGATATATAGTCCACAGAGGGGATATAGAGCTCCACAGAGATAT 635

QY 601 GAGGCCCAACAGAGGATATAGAGCCCACTTATGAGATATAGAGCCGCTGTGGA 660

DB 636 GAGGCCCAACAGAGGATATAGAGCCCACTTATGAGATATAGAGCCGCTGTGGA 695

QY 661 TATGAGTCCCACTTGTGGGATATAGAGCCCACTTGTGGGATATAGAGCCGCTGTGGA 720

DB 696 TATGAGTCCCACTTGTGGGATATAGAGCCCACTTGTGGGATATAGAGCCGCTGTGGA 755

QY 721 GATATATGAGCCCACTTGTGGGATATAGAGCCCACTTGTGGGATATAGAGCCGCTGTGGA 780

DB 756 GATATATGAGCCCACTTGTGGGATATAGAGCCCACTTGTGGGATATAGAGCCGCTGTGGA 815

QY 781 GGGGATATGAGCCCACTTGTGGGATATAGAGCCCACTTGTGGGATATAGAGCCGCTGTGGA 840

DB 816 GGGGATATGAGCCCACTTGTGGGATATAGAGCCCACTTGTGGGATATAGAGCCGCTGTGGA 875

QY 841 CCCCCTCATATGAGGCTTGTGGAATATAGAGCCCACTTGTGGAATATAGAGCCGCTGTGGA 898

DB 876 CCCCCTCATATGAGGCTTGTGGAATATAGAGCCCACTTGTGGAATATAGAGCCGCTGTGGA 933

RESULT 2

ID AAS20602 standard; CDNA; 1001 BP.

XX AC AAS20602;

XX XX 09-APR-2002 (first entry)

XX DT

XX XX DNA encoding human testicular WW domain binding protein (hewbp).

XX DB

XX XX Testicular WW domain binding protein; WBPH; perinuclear theca 32;

XX KW PT32; contraceptive; fertility; oocyte activation; vaccine;

XX KW globozoospermy; spermiogenesis; tyrosine kinase; c-Yes;

XX KW immunoprecipitative; human; gene; ss.

XX OS Homo sapiens.

XX XX

XX PH Key Location/Qualifiers

XX FT 1..18

XX FT primer_bind /note= "Binds forward primer AAS20605"

XX FT CDS 1..708

XX FT /tag= b

XX FT /product= "hewbp"

XX FT /note= "Human testicular WW domain binding protein"

XX FT primer_bind complement (984..1001)

XX FT /tag= c

XX FT /note= "Binds reverse primer AAS20606"

XX FN WO200190185-A2.

XX PD 29-NOV-2001.
 XX PF 25-MAY-2001; 2001MO-CA00738.
 XX PR 25-MAY-2000; 2000CA-2207128.
 XX PR 25-MAY-2000; 2000US-206979P.
 XX PA (TOOH) UNIV QUEBENS KINGSTON.
 XX PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX PI Olo R, Sutovsky P;
 XX DR WPI; 2002-097644/13.
 XX DR P-PSDB; AAU74610.
 PT Isolated perinuclear theca 32 polypeptide that interacts with activated
 PT tyrosine kinase c-Yes, for enhancing fertility, creating/diagnosing
 PT diminished fertility and abnormal spermiogenesis and for providing
 PT contraception -
 XX
 XX PS Claim 62; Fig 4B; 103pp; English.
 CC The invention describes an isolated perinuclear theca 32 (PT32)
 CC polypeptide (1) which interacts with tyrosine kinase c-Yes. (1) is
 CC useful for: enhancing fertility in a mammal; treating globozoosperm, by
 CC expressing (1) in spermatozoa; inhibiting fertilisation, by introducing
 CC (1) or its antigenic fragment into a mammal to elicit an immune
 CC response; enhancing the ability of round spermatids to activate oocytes;
 CC treating or diagnosing diminished fertility and abnormal spermiogenesis;
 CC in providing contraception; identifying contraceptive and
 CC fertility-enhancing agents. The polynucleotide is useful for producing
 CC (1) by recombinant techniques, as vaccine, as diagnostic reagents, and
 CC for chromosome identification. An antibody against (1) is useful in
 CC immunological assays, in immunoneutralization methods, to identify cells
 CC expressing (1), and to purify (1) by affinity chromatography. A
 CC transgenic animal is useful as an animal model for studying human
 CC fertility and reproductive biology, and for screening compounds to
 CC identify modulators of oocyte activation. The use of (1) prevents the
 CC entry of components which are detrimental to embryonic development into
 CC the oocyte during oocyte activation with crude sperm extract and avoids
 CC the propagation of viruses such as HIV (human immunodeficiency virus) and
 CC SIV (simian immunodeficiency virus) carried in the sperm. This sequence
 CC encodes the human testicular MW domain binding protein (hMBP), described
 CC in the method of the invention.
 XX
 XX SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
 Query Match 41.7%; Score 374.4; DB 24; Length 1001;
 Best Local Similarity 75.3%; Pred. No. 5,1e-101;
 Matches 497; Conservative 0; Mismatches 151; Indels 12; Gaps 2;
 QY 223 ATGCGCTTGGCTGATGAGTGAAGTGCACCATTTGAAACAATTTTCCGCCCACTAC 282
 DB 1 ATGCGCATTTGATCTGATGAGTGAAGTGCACCATTTGAAACAATTTTCCGCCCACTTC 60
 QY 283 ATTAAGGAACCATTTAGAGGAGTGCAGAGTGTGCTGGGAGAGAGCAACTGTTTAAAG 342
 DB 61 ATTAAGGAACCATTTAGAGGAGTGCAGAGTGTGCTGGGAGAGAGCAACTCTTTTAA 120
 QY 343 TTATCCCTTCAGGAAAGAGAGTGCATGAAATTTGCCCACTGATGTAAGGTGCTCT 402
 DB 121 TTATGCTTCAGAAATGGAATGCAATGCAATTTGCTCCAGTTGATGTAAGGTGCTCT 180
 QY 403 GCTGCTGCGAGAGAAATTCACCTTGGAATGTAAATTAATCTGTTCCAGCACTTCAGAGCTG 462
 DB 181 GCTGCTGCGCGAGAAATTCACCTTAATTAATGATGCTGTTGAGCTCTTAAGGAAT 240
 QY 463 TACATTAATTAATCTGCTCCAGAGGCTGAGAGTGTCTCTCAGAGCAACTTGTTCAGATAT 522
 DB 241 TATGTAATTAATCTGCGAGAGGAAT--ATGTGACATCCCAAGATGCTGTGTTCAG----- 292
 QY 523 CCAATGTGATCTATGAGACCCCAAGCAAGATATATAGTCAACCAAGGAGAAATATGGA 582

DB 293 ----TTATGTCTATGAGGAGCCCACTGAGATATGAGGCCCACTCCCGAATACGA 348
 QY 583 ACTCCACGAGAGGATATGAGAGCCCAACAGGGGATATGAGGCCCACTATGAGATAT 642
 DB 349 GCTCCACCTGCAAGATATGAGAGCCCAACCTGATGAAATGAGAGGCCCTGTGAGATAC 408
 QY 643 GGAGCCCGCTGTGAGATATGAGATCCCACTGGAGGATATGAGATCCCACTGGAGGA 702
 DB 409 AAGGCTCACTGTGAGATATGAGAGCCCACTCTTGTGATACGAGGCCCACTTCAGGA 468
 QY 703 TATGAGTCCCACTGGAGATATGAGAGCCCACTGGAGGATATGAGATCCCACTGAG 762
 DB 469 TATGAGAGCCCACTCTGAGATATGAGAGCCCACTCTGATATGAGAGCCCACTGTC 528
 QY 763 GGATATGAGAGCCCACTGGAGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTG 822
 DB 529 GGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTCTGAGATATGAGAGCCCACTG 588
 QY 823 GCTGGAATGAGAGCCCTACCTCCCTGATATGAGATCTGCTGGAATATACAGTGC 882
 DB 589 GCGGATATGAGAGCCCTACCTCTGATATGAGAGCCCACTGAGATATGAGAGCC 648
 RESULT 3
 AAL01229
 ID AAL01229 standard; cDNA; 467 BP.
 XX
 XX AAL01229;
 AC
 AC 21-NOV-2001 (first entry)
 DT
 DT Human reproductive system related antigen cDNA SEQ ID NO: 1230.
 XX
 XX Human, reproductive system related antigen; reproductive system disorder;
 XX cancer; gene therapy, se.
 XX
 XX Homo sapiens.
 XX
 XX WO200155320-A2.
 PD 02-AUG-2001.
 XX
 XX 17-JAN-2001; 2001MO-US01339.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-021647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225265.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 06-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI, 2001-46570/50.
 DR P-PSDB, AAM95259.
 DR
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen
 XX
 PT is used in preventing, treating or ameliorating a medical condition
 XX
 PS Claim 1; SEQ ID NO 1230; 1297bp + Sequence listing; English.
 CC
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a coding sequence of the
 CC invention.
 CC
 XX
 SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
 Query Match 29.2%; Score 262.6; DB 2; Length 467;
 Best Local Similarity 76.4%; Pred. No. 7.2e-68;
 Matches 365; Conservative 0; Mismatches 100; Indels 13; Gaps 3;
 QY 351 CAGAAAGAGGTCGTCATGCAATTTGCCCACTGATGTAAGCTGCTCTGCTGTC 410
 DB 2 CAGAAATGAGGTCATGTAATTTGCCCACTGATGTAAGCTGCTCTGCTGTC 61
 QY 411 CAGAGGATTCGACTTGGAAGTGAATTAATTAAGTGGTTCGACCTTGAGACTGTACATTAAT 470
 DB 62 CGAGGATTTCCACTTGAATTAATTAATTAAGTGGTTCGACCTTGAGACTGTACATTAAT 121
 QY 471 TACTGTCCAGGGGCTGCAAGTGTGCTCTGACAGACCTTGTCCAGCATATCAATGT 530
 DB 122 TACTGGGGAAGGAAT---ATGTGCACTCAGAGATGCTTGTTCAG-----TTAT 169
 QY 531 GATCTATGACCCCGACACGACGATATACAGTCAACGAGGGAATATGAACTGCACC 590

DB 170 TGTCTATGAGCCCGCCACTGACAGATATGAGGCCCGCCCGGATAGAGGCCCGCCAC 229
QY 591 AGAAGATATGAGGCCCGCCACAGAGGATATGAGGCCCGCCCGATATGAGGCCCGCC 650
DB 230 TGCAGATATGAGGCCCGCCCGCTAGAAATGAAAGCCCGCTGTGGATACAGAGCCCTC 289
QY 651 GCTGTGAGATATGAGGTCCACCTGAGGGGATATGAGTCCCGCCCGGAGATATGAGCT 710
DB 290 ACTGTGCAATATGAGGCCCGCCCGCTGTGATGAGAGGCCCGCCCGTGCAGATATGAGC 349
QY 711 CCCACCTGAGGAGATATGAGGCCCGCCACCTGAGGGGATATGAGTCCCGCTGAGGGATATG 769
DB 350 CCCACCTGAGGATATGAGGCCCGCCCGCTGTGATGAGAGGCCCGCCCGTGTGAGATG 409
QY 770 GTGCCCCACCTGAGGAGATATGAGGCCCGCCCGCTGAGGATATGAGGCCCGCCCGCTG 827
DB 410 GAGCCCCACCTGAGGATATGAGGCCCGCCCGCTGAGGATATGAGGCCCGCCCGCTG 467

RESULT 4

ABL96688

ID ABL96688 standard; cDNA; 467 BP.

XX ABL96688;

XX 21-JUN-2002 (first entry)

DE Human testicular antigen encoding cDNA SEQ ID NO: 356.

XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;

XX reproductive system disorder; urinary system disorder; gene therapy;

XX cardiovascular disorder; respiratory disorder; neurological disorder;

XX gastrointestinal disease; infection; cytostatic; gene; ss.

OS Homo sapiens.

XX WO200155317-A2.

XX 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01329.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216447.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227059.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0235327.
PR 29-SEP-2000; 2000US-0235367.
PR 29-SEP-2000; 2000US-0235368.
PR 29-SEP-2000; 2000US-0235369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0244676.
PR 08-NOV-2000; 2000US-0244677.
PR 08-NOV-2000; 2000US-0244678.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0255676.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483332/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
XX PT useful for preventing, diagnosing and/or treating testicular cancer -
XX PT
XX
XX Claim 1; SEQ ID NO 356; 768bp; English.
XX
XX The present invention provides the protein and coding sequences of 973
XX CC human testicular antigens, and fragments of their genomic sequences. The
XX CC sequences can be used in the treatment of cardiovascular, urinary system,
XX CC reproductive system, immune, respiratory, neurological and
XX CC gastrointestinal disorders, infections, and particularly cancer,
XX CC especially testicular cancers. The present sequence is a cDNA of the
XX CC invention.
XX
XX Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
SQ
Query Match 29 2%; Score 262.6; DB 23; Length 467;
Best Local Similarity 76.4%; Pred. No. 7.2e-68;
Matches 365; Conservative 0; Mismatches 100; Indels 13; Gaps 3;
QY 351 CAGAAAGAGAGTGCATCAATTTGCCCACTGATGTAAAGCTGCTCTGCTGCTGC 410
DB 2 CAGAAATGAGAGTGCATTTGATTTGCCAGTATGTGAAAGCTGCTCTGCTGCTGC 61
QY 411 CAGAGAAATTCATCTGGAAGTGAATTAATTAATTAATTAATTAATTAATTAAT 470
DB 62 CCGAGAAATTCATCTGGAAGTGAATTAATTAATTAATTAATTAATTAATTAAT 121
QY 471 TACTGTCCAGGGGCTGAGTGTGCTCCACAGACACTGTGTCAGATATCAATGT 530
DB 122 TACTGTCCAGGGGCTGAGTGTGCTCCACAGACACTGTGTCAGATATCAATGT 169
QY 531 GATCTATGACCCCGACACAGATATATCAAGTCCACAGGGGAAATATGAAATCCACC 590

DB 170 TGTCTATGAGCCCACTGACAGATATGAGCCCACTCCGATATACGAGCCCAACC 229
QY 591 AGAAGATATGAGCCCACTGACAGATATGAGCCCACTCCGATATACGAGCCCAACC 650
DB 230 TGCAGATATGAGCCCACTGACAGATATGAGCCCACTCCGATATACGAGCCCAACC 289
QY 651 GCCTGTGAGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGT 710
DB 290 ACCGTGTGATATGAGCCCACTGATATGAGCCCACTGATATGAGCCCACTGATATG 349
QY 711 CCCACTGAGGATATGAGCCCACTGAGGATATGAGTCCCACTGAGGATATGAGT 769
DB 350 CCCACTGATATGAGCCCACTGATATGAGCCCACTGATATGAGCCCACTGATATG 409
QY 770 GTGCCCACTGAGGATATGAGCCCACTGACAGATATGAGCCCACTGAGCCCACTG 827
DB 410 GAGCCCACTGATATGAGCCCACTGACAGATATGAGCCCACTGACAGATATGAGCC 467
RESULT 5
ID AAL04882 standard; DNA; 7099 BP.
XX
XX AAL04882;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 7570.
XX
XX Human; reproductive system related antigen; reproductive system disorder;
XX KW cancer; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX WO200155320-A2.
XX
XX
XX 02-AUG-2001.
XX
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0186874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.

Qy 819 ACCAGCTGGAATGAAGCCTACCCCTGCATATGAGCTCCATCTGCGAATATACAGC 878
Db 7018 GCCTGCGGATACAGAGCCTCACTGCTGATCAAGAGCCAGAGCCTCAGGAATCTACAGC 7077
Qy 879 TGCC 882
Db 7078 AGCC 7081

RESULT 6
ABL97776
ID ABL97776 standard; DNA; 7099 BP.
XX ABL97776;
AC
XX 21-JUN-2002 (first entry)
DT
XX
DE Human testicular antigen encoding DNA fragment SEQ ID NO: 2428.
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic; gene; de.
XX
OS Homo sapiens.
XX
FN WO200155317-A2.
XX
PD 02-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US01329.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205155.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0254779.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483332/52.
XX
XX Nucleic acids encoding 973 human testicular antigen polypeptides,
XX useful for preventing, diagnosing and/or treating testicular cancer -
XX
XX Disclosure; SEQ ID NO 2428; 766bp; English.
XX
XX The present invention provides the protein and coding sequences of 973
XX human testicular antigens, and fragments of their genomic sequences. The
XX sequences can be used in the treatment of cardiovascular, urinary system,
XX reproductive system, immune, respiratory, neurological and
XX gastrointestinal disorders, infections, and particularly cancer.
XX especially testicular cancers. The present sequence is a DNA encoding a
XX protein fragment of the invention.
XX
XX Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
SQ
Query Match 24.5%; Score 220; DB 23; Length 7099;
Best Local Similarity 75.3%; Pred. No. 1.4e-54;
Matches 274; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 519 ATATCCATTGTGATCTATGAGCCGCCCAACAGATATACATCCCAACGAGGGAATA 578
DB 6718 ATTCCTCATATGTCTATGAGCCGCCCACTGAGATATGAGCCCACTCCGAGATA 6777
QY 579 TGAACCTCCACAGAGATATGAGCCCAACAGAGGATATGAGCCCACTATGAGG 638
DB 6778 CGAGACCCCACTGAGATATGAGCCCAACCTAGAGATATGAGCCCGCTGTGGG 6837
QY 639 ATATGAGACCCCGCTGTGAGATATGAGTCCCACTTGGGAGATATGAGTCCCACTGG 698
DB 6838 ATACGAGACCTTCACCTGTGAGATATGAGCCCACTTGTGATATGAGACCCCACTGCG 6897
QY 699 GGGATATGAGATCCCACTGAGGAGATATGAGCCCACTTGGGAGATATGAGATCCCACTG 758
DB 6898 AGGATATGAGACCCCACTTCATGAGATATGAGCCCACTTGTGATATGAGACCCCACTG 6957
QY 759 TGGGAGATATGAGTCCCACTGAGGAGATATGAGCCCACTTCATGAGATATGAGACCC 818
DB 6958 TCTGAGATATGAGACCCCACTTCCTGAGATATGAGCCCACTTCGAGAGATATGAGACCC 7017
QY 819 ACCAGCTGAAATGAGACCCCTACCCCTGATATGAGAGCTTCATCTGTGAGAAATACAGC 878
DB 7018 GCTGCGGAGATACAGAGCTTCACTGTGAGATACAGAGAGGCTTCAGAGATATACAGC 7077

QY 879 TGCC 882
DB 7078 AGCC 7081
RESULT 7
ABK84798
ID ABK84798 standard, cDNA; 220895 BP.
XX
XX ABK84798;
AC
XX 14-AUG-2002 (first entry)
DT
XX
XX Human cDNA differentially expressed in granulocytic cells #1369.
DE
XX
XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
XX viral infection; parasitic infection; protozoal infection;
XX fungal infection; sterile inflammatory disease; psoriasis;
XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
XX cardiac reperfusion injury; renal reperfusion injury; AIDS;
XX adult respiratory distress syndrome; inflammatory bowel disease;
XX Crohn's disease; ulcerative colitis; periodontal disease;
XX granulocyte activation; chronic inflammation; allergy.
XX
XX Homo sapiens.
OS
XX
XX WO200228999-A2.
PN
XX
XX 11-APR-2002.
PD
XX
XX 03-OCT-2001; 2001WO-US30821.
PF
XX
XX 03-OCT-2000; 2000US-237189P.
PR
XX
XX (GENE-) GENE LOGIC INC.
PA
XX
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
PI WPI; 2002-435328/46.
XX
XX
XX Detecting granulocyte activation by detecting differential expression
XX of genes associated with granulocyte activation, which serves as
XX diagnostic markers that is useful for monitoring disease states and
XX drug toxicity -
XX
XX Claim 1; SEQ ID NO 1369; 114bp; English.
PS
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing
XX the expression level to an expression level in an unactivated
XX GC, where differential expression (M2) Gs by contacting GC with an agent
XX that alters the expression of at least one gene in Gs; (2) screening (M3)
XX for an agent capable of modulating GCA or an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease using the
XX gene expression profile; (3) detecting (M4) an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease, by detecting the
XX level of expression in a sample of the tissue of gene(s) from Gs, where
XX the level of expression of the gene is indicative of inflammation;
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,
XX or allergic response in a subject, exposure of a subject to a pathogen
XX or sterile inflammatory disease, by contacting a tissue having
XX inflammation with an agent that modulates the expression of gene(s)
XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
XX modulating Gs; M3 is useful for screening an agent capable of modulating
XX GCA preferentially in an inflammation (especially chronic) in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal

CC reperfusion injury, ARDS, adult respiratory distress syndrome,
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC periodontal disease; also bacterial infection, viral infection,
CC paratubercular infection, protozoal infection, fungal infection and MS is
CC useful for treating one of the above conditions. The present
CC sequence represents a gene differentially expressed in granulocytes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 220895 BP; 53760 A; 54808 C; 55831 G; 56496 T; 0 other;
Query Match 24.5%; Score 220; DB 24; Length 220895;
Best Local Similarity 75.3%; Pred. No. 8.4e-54;
Matches 274; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 519 ATATCCAAATGATGATCTATGAGACCCCAACAGAGATATACATCCACAGGGGATA 578
DB 169998 ATTCCAGTTATGTTCTATGAGACCCCACTGAGATATGAGACCCCACTCCGGATA 170057
QY 579 TGGAACTCCACAGAGATATGAGACCCCAACAGGGGATATGAGACCCCACTTATGG 638
DB 170058 CGGAGCCCACTGAGAGATATGAGACCCCAACAGGGGATATGAGACCCCACTTATGG 170117
QY 639 ATATGAGACCCCACTGAGAGATATGAGACCCCACTGAGAGATATGAGACCCCACTTATGG 698
DB 170118 ATACAGAGCTTCACTGAGATATGAGACCCCACTTATGATATGAGACCCCACTTATGG 170177
QY 699 GGGATATGAGTCCCACTGAGAGATATGAGACCCCACTGAGAGATATGAGACCCCACTTATGG 758
DB 170178 AGGATATGAGACCCCACTGAGAGATATGAGACCCCACTTATGATATGAGACCCCACTTATGG 170237
QY 759 TGGGAGATATGAGTCCCACTGAGAGATATGAGACCCCACTGAGAGATATGAGACCCCACTTATGG 818
DB 170238 TCTGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGG 170297
QY 819 ACCGATGAGAAATGAGACCCCACTGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGG 878
DB 170298 GCGTGGGAGATGAGACCCCACTGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGG 170357
QY 879 TGCC 882
DB 170358 AGCC 170361
RESULT 8
ABA67850
ID ABA67850 standard; DNA; 436 BP.
XX
AC ABA67850;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #16155.
XX
KM Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

XX
PA (MOLR-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI, 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human foetal liver -
XX
PS Claim 4; SEQ ID NO 16155; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
Query Match 24.1%; Score 216.8; DB 22; Length 436;
Best Local Similarity 75.6%; Pred. No. 3.1e-54;
Matches 269; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 527 TTGTGATCTATGAGACCCCAACAGAGATATGAGACCCCACTTATGAGATATGAGAACTC 586
DB 1 TTATGATCTATGAGACCCCACTGAGAGATATGAGACCCCACTTATGAGATATGAGAACTC 60
QY 587 CACGAGAGATATGAGACCCCAACAGAGATATGAGACCCCACTTATGAGATATGAGAACTC 646
DB 61 CACCTGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGAACTC 120
QY 647 CCCGCTGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGAACTC 706
DB 121 CTTACCTGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGAACTC 180
QY 707 GAGTCCCACTGAGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGAACTC 766
DB 181 GAGCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGAACTC 826
QY 767 ATGATCCCACTGAGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGAACTC 882
DB 241 ATGAGCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGAACTC 940
QY 827 GAAATGAGCCCTTATGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGAACTC 882
DB 301 GATTCAGAGCTTATGAGATATGAGACCCCACTTATGAGATATGAGACCCCACTTATGAGATATGAGAACTC 940
RESULT 9
AAK42003
ID AAK42003 standard; DNA; 436 BP.
XX
AC AAK42003;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 16560.
XX
KM Human; bone marrow expressed exon; gene expression analysis; probe;
KM microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.


```

XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO: 16560; 658bp + sequence listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is one of
XX the probes of the invention.
XX
SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
Query Match 24.1%; Score 216.8; DB 22; Length 436;
Best Local Similarity 75.6%; Pred. No. 3.1e-54;
Matches 269; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 527 TTGTATCTATGAGACCCCAACGAGATATACAGTCCAAACGAGGAAATATGAACTC 586
DB 1 TTATGTCTATGAGACCCCAACGAGATATGAGACCCCAACCTCCGATTCGAGGCC 60
QY 587 CACGAGAAGATATGAGACCCCAACGAGGAGATATGAGACCCCAACCTATGAGATATGAG 646
DB 61 CACTGCAAGATATGAGACCCCAACCTGAGAAATATGAGACCCCAACCTGAGATACAGAG 120
QY 647 CCCCCTGTGAGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGAGATATG 706
DB 121 CCTCACTGTGAGATATGAGACCCCACTCTTGATATGAGACCCCACTGAGAGATATG 180
QY 707 GAGTCCCACTGAGGATATGAGTCCCACTGAGGAGATATGAGTCCCACTGAGGAGAT 766
DB 181 GAGCCCACTCTAGGATATGAGACCCCACTCTTGATATGAGACCCCACTCTCGAGAT 240
QY 767 ATGTGCCCCCACTGAGGAGATATGAGACCCCACTGAGGAGATATGAGACCCCACTG 826
DB 241 ATGTGAGCCCACTCTGAGATATGAGACCCCACTGAGGAGATATGAGACCCCACTG 300
QY 827 GAAATGAAGCCCTACCCCTGATATGAGTCCATCTCTGAAATATACAGCTGCC 882
DB 301 GATACAGAGCTCACTCTGATATGAGACCCCACTGAGGAGATATGAGACCCCACTG 356
XX
RESULT 10
AA148070
ID AA148070 standard; DNA; 436 BP.
XX
XX AA148070;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #16756 used to measure gene expression in human placenta sample.
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; 88.
XX
XX Homo sapiens.
XX

```

```

PN W0200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001MO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta.
XX
XX Claim 25; SEQ ID No 16756; 654bp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX The present sequence is one such probe. The probes are useful for
XX producing a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
Query Match 24.1%; Score 216.8; DB 22; Length 436;
Best Local Similarity 75.6%; Pred. No. 3.1e-54;
Matches 269; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 527 TTGTATCTATGAGACCCCAACGAGATATACAGTCCAAACGAGGAAATATGAACTC 586
DB 1 TTATGTCTATGAGACCCCAACCTGAGAAATATGAGACCCCAACCTCCGATTCGAGGCC 60
QY 587 CACGAGAAGATATGAGACCCCAACGAGGAGATATGAGACCCCAACCTATGAGATATGAG 646
DB 61 CACTGCAAGATATGAGACCCCAACCTGAGAAATATGAGACCCCAACCTGAGATACAGAG 120
QY 647 CCCCCTGTGAGATATGAGTCCCACTGAGGAGATATGAGTCCCACTGAGGAGATATG 706
DB 121 CCTCACTGTGAGATATGAGACCCCACTCTTGATATGAGACCCCACTGAGAGATATG 180
QY 707 GAGTCCCACTGAGGAGATATGAGTCCCACTGAGGAGATATGAGTCCCACTGAGGAGAT 766
DB 181 GAGCCCACTCTAGGATATGAGACCCCACTCTTGATATGAGACCCCACTCTCGAGAT 240
QY 767 ATGTGCCCCCACTGAGGAGATATGAGACCCCACTGAGGAGATATGAGACCCCACTG 826
DB 241 ATGTGAGCCCACTCTGAGATATGAGACCCCACTGAGGAGATATGAGACCCCACTG 300
QY 827 GAAATGAAGCCCTACCCCTGATATGAGTCCATCTCTGAAATATACAGCTGCC 882
DB 301 GATACAGAGCTCACTCTGATATGAGACCCCACTGAGGAGATATGAGACCCCACTG 356
XX
RESULT 11
ABS16034
ID ABS16034 standard; DNA; 436 BP.
XX
XX ABS16034;
XX
XX 19-AUG-2002 (first entry)
XX
XX Human genome-derived single exon probe ORF from lung SEQ ID No 16025.
XX
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
XX

```

KM chronic obstructive pulmonary disease; interstitial lung disease;
 KM familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KM tuberculous scleriosis; Gaucher's disease; Niemann-Pick disease;
 KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemostidrosis;
 KM pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;
 KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KM primary ciliary dyskinesia; pulmonary hypertension;
 KM hyaline membrane disease; open reading frame; ORF.
 XX
 OS Homo sapiens.
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 DR
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 PT
 XX
 PS Claim 4; SEQ ID No 16025; 634bp; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridize at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridization of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridization to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probe/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberculous scleriosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemostidrosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Kargener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a single exon
 CC probe open reading frame of the invention.
 CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;
 Query Match 24.1%; Score 216.8; DB 24; Length 436;
 Best Local Similarity 75.6%; Pred. No. 3.1e-54;
 Matches 269; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 527 TTGTGATCTATGAGACCCCAACAGATATACAGTCCAAAGGGGAATATGAACTC 586
 DB 1 TTATGTCTATGAGACCCCACTGACGATATGAGACCCCACTCCGGATACGAGCCC 60
 QY 587 CACCAAGAGATATGAGACCCCAACGAGGGGATATGAGACCCCACTATAGGATATGAG 646
 DB 61 CACCTCAGATATGAGACCCCAACCGTGAAGAAATGAAGCCCGCTGTGGATACAGAG 120
 QY 647 CCCCGCTGTGGATATGAGTCCCACTGAGGGATATGAGTCCCACTGAGGGATATG 706
 DB 121 CTCACCTGTGCAATATGAGACCCCACTCTTGAATACGAGCCCACTGCAAGATATG 180
 QY 707 GAGTCCCACTGAGGGATATGAGACCCCACTGAGGGATATGAGTCCCACTGAGGGAT 766
 DB 181 GAGCCCACTTACGATATGAGACCCCACTCTTGAATATGAGACCCCACTCTGGAT 240
 QY 767 ATGTGCCCCCACTGAGGGATATGAGACCCCACTCTGCAAGATATGAGACCCCACTG 826
 DB 241 ATGAGGCCCACTCTGAGATATGAGACCCCACTTCAAGAAATGAAGCCCGCTGCGG 300
 QY 827 GAATGAAGCCCTACCCCTGCAATATGAGCTCATCTGTGGAATATACAGCTGCC 882
 DB 301 GATACAGGCTTCACTGCTGATACAGAGCCAGGCTTCAAGAAATATACAGCAGCC 356
 RESULT 12
 ID AAL01355 standard; cDNA; 321 BP.
 XX
 AC AAL01355;
 XX
 DT 21-NOV-2001 (first entry)
 XX
 DB Human reproductive system related antigen cDNA SEQ ID NO: 1356.
 XX
 KM Human; reproductive system related antigen; reproductive system disorder;
 KM cancer; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200155320-A2.
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01339.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234297.
PR 25-SEP-2000; 2000US-0234298.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0255719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
DR P-PSDB; AAM95385.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; SEQ ID NO 1356; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
CC invention.
XX
SQ Sequence 321 BP; 83 A; 75 C; 70 G; 89 T; 4 other;

Query Match 21.8% Score 195.4; DB 22; Length 321;
Best Local Similarity 77.4%; Pred. No. 6.3e-48;
Matches 246; Conservative 2; Mismatches 69; Indels 1; Gaps 1;
QY GCGCTGCGGGGCGCTCATCCCTTGGCGAAAGTGTCTTGAAGCAGGTGAGATGTCG 88
|||||

PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM,
 XX
 DR WPI: 2001-48332/52.
 XX
 PT Nucleic acids encoding 973 human testicular antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating testicular cancer -
 XX
 PS Claim 1; SEQ ID NO 476; 766bp; English.
 XX
 CC The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a cDNA of the
 CC invention.
 XX
 SQ Sequence 321 BP; 83 A; 75 C; 70 G; 89 T; 4 other;

Query Match 21.8%; Score 195.4; DB 23; Length 321;
 Best Local Similarity 77.4%; Pred. No. 6.3e-48;
 Matches 246; Conservative 2; Mismatches 69; Indels 1; Gaps 1;
 QY 29 GCGCTGCGGCGCCCTCATCCCTCTGCGGAAAGTCTTGAAGCAAGTGAAGATGTGG 88
 DB 1 GCGCGCGGCGGCGCCCTCATCCCTCAAGCGTGAAGCTCTTGAAGCGGTCTCCGAATGTGG 60
 QY 89 ACCTCTGCTTCTTCAAGAAACCAAGTGAATCTTATCTTTATGCAAGAAAGAA 148
 DB 61 AGCTCTCTTCCCAAGCAAGCAAGGCTCAATGTCTTTAGTGAAGAAAGCAAGAA 120
 QY 149 CGTGTGTCACCTTATACCGGGGAGTGTGTGTCGACCTTCAAGTCAATGACCCCA 208
 DB 121 CATGTTCTCACTTATACCGGGGAGTGTGTGTCGACCTTCAAGTCAATGACCCCA 180

QY 209 TGTCTTCTTATGATGCGTTTGACCTGATGATGATGATGATGATGATGATGATGAT 268
 DB 181 TGTCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 QY 269 TTGCCCCCACTTACATTAAGAAACATTCAGGAGCTCCAGGTGTGCTGG-AMGA 327
 DB 241 TTGCTGCAACTTATTAAGGAACTATTGAGGAGCTTCATATGAGGCTGGAAAGGA 300
 QY 328 CAAGCTGTTTAAAGTTA 345
 DB 301 CAAGCTCTTTAAANTA 318

RESULT 14

ABST0430
 ID ABST0430 standard; cDNA; 1885 BP.
 XX
 XX ABST0430;
 AC
 XX 27-NOV-2002 (first entry)
 DT
 XX Human bone remodelling gene #87.
 XX
 XX Bone remodelling; osteoporosis; human; gene; ss.
 XX
 XX Homo sapiens.
 OS
 XX US6426186-B1.
 PN
 XX 30-JUL-2002.
 PD
 PF 18-JAN-2000; 2000US-0484970.
 XX
 XX 18-JAN-2000; 2000US-0484970.
 PR
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Jones KA, Volkmutz W, Walker MG;
 PI
 XX WPI: 2002-673014/72.
 DR
 XX
 XX A combination of polynucleotides which are co-expressed with genes
 PT known to be involved in bone remodeling and osteoporosis are useful in
 PT an array for the diagnosis of bone remodeling and osteoporosis
 PT associated disorders -
 XX
 PS Claim 1; Column 247-250; 206bp; English.
 XX
 CC The invention relates to a combination comprising a number of
 CC substantially purified and isolated polynucleotides which are
 CC co-expressed with genes known to be involved in bone remodeling and
 CC osteoporosis. The invention is used to diagnose disorders associated
 CC with bone remodeling or osteoporosis. ABST0344-ABST0512 represent
 CC human bone remodeling genes of the invention.
 XX
 SQ Sequence 1885 BP; 384 A; 610 C; 476 G; 415 T; 0 other;

Query Match 13.6%; Score 122; DB 24; Length 1885;
 Best Local Similarity 59.4%; Pred. No. 1.2e-25;
 Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;
 QY 1 ATGCAAGTGAACCAAGCCACACGAGAGCGGTGCGGAGCCCTCATCCCTCTGCGGAA 60
 DB 61 ATGCGCTCAACAAAGAAATCACTCGAGGCGCGCGAGTGAATGTCATTAACAC---CGAG 117
 QY 61 AGTGTCTTGAAGCAAGTGAAGATGGAACCTGTGCTTCTTACAGAAACAGTGAATCC 120
 DB 118 ACCATCTTAATTTCTTATGATCACTGGAACCTCAATTCATATGACATGAAGACCTGCCA 177
 QY 121 TATCTTTAATGACAAAGAAAGAAAGTGTCTTCACTTCACTTACCGGGTGTCTTC 180
 DB 178 GAAGCTTCAAGGAGCAAGAAAGCAAGTGTCTTCACTTCACTTACCGGGTGTCTTC 237

QY 181 GTGACTTCACACTTGAATGACCCGAGCTTTCTTTATGATGCGCTTGACGTGAG 240
 DB 238 CTGTC-----CAAGGGCAAGATGCCATGACGCTTCAATGATGCAATTTATTCATG 291
 QY 241 AGTACTGACCACTTGAACCAACCAATTTTCCGCCCACTACATTAAGAACCATTCAG 300
 DB 292 AAGACTGTGATGATCAAGACGAGCCGCTATTTGTGCAAACTACATCAAGGAAACAGTGAAG 351
 QY 301 GCAGCTCCAGTGTGCTGGGAGAGCAAGCTGTTTAAATGATTCCTTCAGAAAGGA 360
 DB 352 GCGAAGCGGAGAGTGGCTGGGAAGCTCTGCTTCCACAAATTTGACTTTCAGCGGAGG 411
 QY 361 GTGCGCATGCAATTTGCGCACTGATGTAAGTGAAGCTGCTGCTGCTGCGAGAG 416
 DB 412 GCGCGCATTTGATTTGCGAAGCGGATGCTCGAGTGGCATCTCAAGCTTCAGAGG 467

RESULT 15

AAF22363
 ID AAF22363 standard; cDNA; 1915 BP.

AC AAF22363;

DT 26-MAR-2001 (first entry)

DE Human secreted protein gene 48 SEQ ID NO:58.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 XX cerebroprotective; nocotropic; neuroprotective; antibacterial; virucide;
 XX fungicide; ophthalmological; vulnery; gene therapy; neoplasm;
 XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 XX cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 XX cerebral ischaemia; angiogenesis; nervous system disorder; infection;
 XX Alzheimer's disease; ocular disorder; corneal infection; wound healing;
 XX skin aging; food additive; preservative; ss.

OS Homo sapiens.

PN WO200061748-A1.

PD 19-OCT-2000.

PF 06-APR-2000; 2000MO-US08982.

XX 09-APR-1999; 99US-0128696.

PR 14-JAN-2000; 2000US-0176069.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Ruben SM, Komatsoulis G;

PI WPI; 2000-638566/61.

DR P-PSDB; AAB63096.

XX New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -

XX Claim 1; Page 429-430; 480pp; English.

XX AAF22316 to AAF22363 encode the human secreted proteins given in AAB63049
 CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 CC cerebroprotective; nocotropic; neuroprotective; antibacterial; virucide;
 CC fungicide; ophthalmological; and vulnery. The polynucleotides and
 CC proteins can be used to prevent, treat or ameliorate a medical condition
 CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
 CC sheep. They are also used in diagnosing a pathological condition or

CC susceptibility to a pathological condition. Disorders which are diagnosed
 CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angioneitis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
 CC ocular disorders e.g. corneal infection. The polypeptides can also be
 CC used to aid wound healing and epithelial cell proliferation, to prevent
 CC skin aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities, fat content,
 CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
 CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent
 CC sequences used in the exemplification of the present invention.

XX Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;

Query Match 13.6%; Score 122; DB 21; Length 1915;
 Best Local Similarity 59.4%; Pred. No. 1.2e-25;
 Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;

QY 1 ATGGCAATGACCAAGCCACACCGAGCGCTGCGGCGCTCATCCCTCTGCGGA 60
 DB 59 ATGGCGCTCAACAAGATCACTCGAGGGCGGAGTATGTCATTAACAC--CGAG 115
 QY 61 AGTGTCTTGAAGCAGTGTGAGATGTGACCTCTGTTCTTACAGAAACAGTGAATCC 120
 DB 116 AGATCTCTAATGTCTTATATATCAGCGAATCTCAATTAATGACATGAAGACGTGCCA 175
 QY 121 TATCTCTTAAATGGCAAGAAGAGCGTTGTTCTCACTTACACCGGCTGCTTC 180
 DB 176 GAAGCTTTAAAGGAGCAAGAAAGGACGTGTACTTACCTTACCGGCTATCTTT 235
 QY 181 GTGACTTCACACTTACTGATGATGACCCGATGCTTTCTTTATGATGCGCTTGATG 240
 DB 236 CTGTC-----CAAGGGCAAGATGTCATGCAAGTCTTCATATGTCATTTATTCATG 289
 QY 241 AGTACTGACCACTTGAACCAACCAATTTTCCGCCCACTACATTAAGAACCATTCAG 300
 DB 290 AAGACTGTGATGATCAAGACGCGCTATTTGTGCAAACTACATCAAGGAAACAGTGAAG 349
 QY 301 GCAGCTCCAGTGTGCTGGGAGAGCAAGCTGTTTAAATGATTCCTTCAGAAAGGA 360
 DB 350 GCGAAGCGGAGAGTGGCTGGGAAGCTGCTCTTCAAGATTACTTTCAGCGAGAG 409
 QY 361 GTGCGCATGCAATTTGCGCACTGATGTAAGTGAAGCTGCTGCTGCTGCGAGAG 416
 DB 410 GCGCGCATTTGATTTGCGAAGCGGATGCTCGAGTGGCATCTCAAGCTTCAGAGG 465

Search completed: December 16, 2003, 11:00:47
 Job time : 241.646 secs

QY 361 GCGCCATGCAATTTTGCCCACTGATGTAAGCTGCTGCTGCTGCAAGG 416
DB 412 GCGCCATGATGTTGCAAGAGGATGCTCCAGGTGATCTCAAGCTCCAGAG 467

RESULT 2

US-08-728-323A-1
Sequence 1, Application US/08728323A

Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Rusco, James J.
APPLICANT: Edelman, Isidore S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 6.7%; Score 59.8; DB 2; Length 3489;
Best Local Similarity 46.9%; Pred. No. 5.9e-08;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 479 CAGGGGCTGAGTGTGCTCTCAAGACACTTGTCAGCATATCCAAATGTGATCTATG 538
DB 2185 CAGCAGCAGCAGATGACAGCAGCAGCAGATGACAGCAGCAGCAGATGACAGCAG 2244
QY 539 GACCCCAACCAAGATATACAGTCCAAACAGGGGATATGAACTCCACAGAGAT 598
DB 2245 CAGCAGATGAAACAGAGCAGCAGAGAGAGCAGAGCAGAGAGCAGAGCAGAG 2304
QY 599 ATGAGCCCAACAGAGGAGATATGAGCCCACTATGGATATGAGCCCTGCTGG 658
DB 2305 TTAGAGGAGCAGAGCAGAGGATTAAGATCAGAGCAGAGGATTAAGAGCAGAGCAG 2364
QY 659 GATATGAGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACTG 718

DB 2365 GAGTTAGAGAGCAGAGCAGAGGATTAAGAGCAGAGCAGAGGATTAAGAGCAGAG 2424
QY 719 GGGGATATGAGCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACT 778
DB 2425 CAGAGTTAGAGAGCAGAGCAGAGGATTAAGAGCAGAGCAGAGGATTAAGAGCAG 2484
QY 779 CTGGGGATATGAGCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACT 838
DB 2485 GAGCAGAGGATTAAGAGCAGAGCAGAGGATTAAGAGCAGAGGATTAAGAGCAGAG 2544
QY 839 TACCCCTGATATGAGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACT 877
DB 2545 CAGAGGTTGAGAGCAGAGCAGAGGATTAAGAGCAGAGGATTAAGAGCAGAG 2583

RESULT 3

US-09-298-568-1
Sequence 1, Application US/09298568

Patent No. 6323792
GENERAL INFORMATION:
APPLICANT: Kieff, Elliott D.
APPLICANT: Ballesca, Mary E.

TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
FILE REFERENCE: 16412-10001R
CURRENT APPLICATION NUMBER: US/09/298,568
CURRENT FILING DATE: 1999-04-21
EARLIER APPLICATION NUMBER: US 60/109,422
EARLIER FILING DATE: 1998-11-19
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

Query Match 6.7%; Score 59.8; DB 4; Length 3489;
Best Local Similarity 46.9%; Pred. No. 5.9e-08;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 479 CAGGGGCTGAGTGTGCTCTCAAGACACTTGTCAGCATATCCAAATGTGATCTATG 538
DB 2185 CAGCAGCAGCAGATGACAGCAGCAGCAGATGACAGCAGCAGCAGATGACAGCAG 2244
QY 539 GACCCCAACCAAGATATACAGTCCAAACAGGGGATATGAACTCCACAGAGAT 598
DB 2245 CAGCAGATGAAACAGAGCAGCAGAGAGAGCAGAGCAGAGCAGAGCAGAGCAGAG 2304
QY 599 ATGAGCCCAACAGAGGAGATATGAGCCCACTATGGATATGAGCCCTGCTGG 658
DB 2305 TTAGAGGAGCAGAGCAGAGGATTAAGATCAGAGCAGAGGATTAAGAGCAGAGCAG 2364
QY 659 GATATGAGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACTG 718
DB 2365 GAGTTAGAGAGCAGAGCAGAGGATTAAGAGCAGAGCAGAGGATTAAGAGCAGAG 2424
QY 719 GGGGATATGAGCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACT 778
DB 2425 CAGAGTTAGAGAGCAGAGCAGAGGATTAAGAGCAGAGCAGAGGATTAAGAGCAGAG 2484
QY 779 CTGGGGATATGAGCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACT 838
DB 2485 GAGCAGAGGATTAAGAGCAGAGCAGAGGATTAAGAGCAGAGGATTAAGAGCAGAG 2544
QY 839 TACCCCTGATATGAGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACT 877
DB 2545 CAGAGGTTGAGAGCAGAGCAGAGGATTAAGAGCAGAGGATTAAGAGCAGAG 2583

RESULT 4

US-09-410-399-1

Sequence 1, Application US/09410399
Patent No. 6482587
GENERAL INFORMATION:
APPLICANT: Robertson, Eric S.
APPLICANT: Cotter, Murray A.
TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
FILE REFERENCE: UM-03778
CURRENT APPLICATION NUMBER: US/09/410,399
CURRENT FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 6
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

Query Match 6.7%; Score 59.8; DB 4; Length 3489;
Best Local Similarity 46.9%; Pred. No. 5.9e-08;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 479 CAGGGCTGCAAGTGTCTCTCAACACACCTTGTCCAGCATTCATTTGATCTATG 538
DB 2185 CAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAG 2244
QY 539 GACCCCAACCAAGATATACAGTCCAAACAGGGGAAATAGAACTCCACCAAGAGAT 598
DB 2245 CAGCAGATGAAACAGACACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304
QY 599 ATGAGCCCAACAGGGGATATAGAGCCCACTATGGATATAGAGCCCGCTGTG 658
DB 2305 TTAGAGAGCAGAGCAGAGAGATTAAGATCAAGAGCAGAGATTAAGAGAGAGAGAG 2364
QY 659 GATATGAGTCCCACTGGGGATATAGAGTCCCACTGGGGATATAGAGTCCCACTG 718
DB 2365 GAGTTAGAGAGCAGAGCAGAGAGATTAAGAGAGAGAGAGAGAGATTAAGAGAGAG 2424
QY 719 GGGGATATGAGCCCACTGGGGATATAGAGTCCCACTGGGGATATAGTGTCCCACT 778
DB 2425 CAGAGATTAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2484
QY 779 CTGGGGATATAGAGCCCACTGAGAGATATAGAGCCCACTGAGAGAGAGAGAGAG 838
DB 2485 GAGCAGAGATTAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2544
QY 839 TACCCCTGCATATGAGCTCCATCTGCTGGAATAACAG 877
DB 2545 CAGAGGTGAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2583

RESULT 5
US-08-770-379-20/c
Sequence 20, Application US/08770379
Patent No. 5849564
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,379
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 52342
TELEPHONE: (212) 391-0525
TELEFAX: (212) 278-0400
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-770-379-20

Query Match 6.7%; Score 59.8; DB 2; Length 32207;
Best Local Similarity 46.9%; Pred. No. 1.7e-07;
Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 479 CAGGGCTGCAAGTGTCTCTCAACACACCTTGTCCAGCATTCATTTGATCTATG 538
DB 19812 CAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAG 19753
QY 539 GACCCCAACCAAGATATACAGTCCAAACAGGGGAAATAGAACTCCACCAAGAGAT 598
DB 19752 CAGCAGATGAAACAGACACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19693
QY 599 ATGAGCCCAACAGGGGATATAGAGCCCACTATGGATATAGAGCCCGCTGTG 658
DB 19692 TTAGAGAGCAGAGCAGAGAGATTAAGATCAAGAGCAGAGATTAAGAGAGAGAGAG 19633
QY 659 GATATGAGTCCCACTGGGGATATAGAGTCCCACTGGGGATATAGAGTCCCACTG 718
DB 19632 GAGTTAGAGAGCAGAGCAGAGAGATTAAGAGAGAGAGAGAGAGATTAAGAGAGAG 19573
QY 719 GGGGATATGAGCCCACTGGGGATATAGAGTCCCACTGGGGATATAGTGTCCCACT 778
DB 19572 CAGAGATTAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19513
QY 779 CTGGGGATATAGAGCCCACTGAGAGATATAGAGCCCACTGAGAGAGAGAGAGAG 838
DB 19512 GAGCAGAGATTAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19453
QY 839 TACCCCTGCATATGAGCTCCATCTGCTGGAATAACAG 877
DB 19452 CAGAGGTGAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19414

RESULT 6
US-08-757-669A-20/c
Sequence 20, Application US/08757669A
Patent No. 6183751
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUR ASSOCIATED KAPOSI'S SARCOMA VIRUS
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York

STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/757,669A
 FILING DATE:
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 45185-P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 32207 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-757-669A-20

Query Match 6.7%; Score 59.8; DB 3; Length 32207;
 Best Local Similarity 46.9%; Pred. No. 1.7e-07;
 Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

479 CAGGGGCTGCAATGTCCTCTCCACAGACCTGTCGAGATATCCAAATGTGATCTATG 538
 DB 19812 CAGCAGCAGCAGATGACGACGACGACGATGACGACGACGACGATGACGACG 19753
 QY 539 GACCCCAACCAAGATATACAGTCCACACAGGGAATATGAACTCCACCAAGAT 598
 DB 19752 CAGCAGATGACAGAGGACGACGACGACGACGACGACGACGACGACGACG 19693
 QY 599 ATGAGCCCAACAGGGGATATGAGCCCACTTATGAGATATGAGCCCGCTGTG 658
 DB 19692 TTAGAGAGCAGAGGAGGAGGATTAAGAGATCAGAGCAGAGGTTAGAGAGCAG 19633
 QY 659 GATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 718
 DB 19632 GAGTTAGAGAGCAGAGCAGGATTAGAGAGCAGAGCAGAGGTTAGAGAGCAG 19573
 QY 719 GGGATATGAGAGCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACT 778
 DB 19572 CAGGAGTTAGAGAGCAGAGGATTAGAGAGCAGAGGATTAGAGAGCAG 19513
 QY 779 CTGGGGATATGAGCCCACTGAGGATATGAGCCCACTGAGGATATGAGCCCACT 838
 DB 19512 GAGCAGAGATTAGAGAGCAGAGGATTAGAGAGCAGAGGATTAGAGAGCAG 19453
 QY 839 TACCCCTGATATGAGCTCCATCTGCTGGAATATCAG 877
 DB 19452 CAGAGGTGAGAGCAGAGCAGAGCAGAGGAAAGCAG 19414

RESULT 7
 US-09-230-371A-20/c
 Sequence 20, Application US/09230371A
 Patent No. 6348586
 GENERAL INFORMATION:
 APPLICANT: Chang, Yuan
 APPLICANT: Bohenzky, Roy A
 APPLICANT: Rusco, James J
 APPLICANT: Beelman, Isidore S
 APPLICANT: Moore, Patrick S
 TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
 TITLE OF INVENTION: US9S THERIOF

FILE REFERENCE: 45185-G-PCT-US
 CURRENT APPLICATION NUMBER: US/09/230,371A
 CURRENT FILING DATE: 1999-11-17
 PRIOR APPLICATION NUMBER: PCT/US97/13346
 PRIOR FILING DATE: 1997-07-22
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: Patent Ver. 2.0
 SEQ ID NO 20
 LENGTH: 32207
 TYPE: DNA
 ORGANISM: Kaposi's sarcoma-associated herpesvirus
 US-09-230-371A-20

Query Match 6.7%; Score 59.8; DB 4; Length 32207;
 Best Local Similarity 46.9%; Pred. No. 1.7e-07;
 Matches 187; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

479 CAGGGGCTGCAATGTCCTCTCCACAGACCTTTCAGAGATATCCAAATGTGATCTATG 538
 DB 19812 CAGCAGCAGCAGATGACGACGACGACGATGACGACGACGACGATGACGACG 19753
 QY 539 GACCCCAACCAAGATATACAGTCCACACAGGGAATATGAACTCCACCAAGAT 598
 DB 19752 CAGCAGATGACAGAGGACGACGACGACGACGACGACGACGACGACGACG 19693
 QY 599 ATGAGCCCAACAGGGGATATGAGCCCACTTATGAGATATGAGCCCGCTGTG 658
 DB 19692 TTAGAGAGCAGAGGAGGAGGATTAGAGAGCAGAGCAGAGGTTAGAGAGCAG 19633
 QY 659 GATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACTG 718
 DB 19632 GAGTTAGAGAGCAGAGCAGGATTAGAGAGCAGAGCAGAGGTTAGAGAGCAG 19573
 QY 719 GGGATATGAGAGCCCACTGAGGATATGAGTCCCACTGAGGATATGAGTCCCACT 778
 DB 19572 CAGGAGTTAGAGAGCAGAGGATTAGAGAGCAGAGGATTAGAGAGCAG 19513
 QY 779 CTGGGGATATGAGCCCACTGAGGATATGAGCCCACTGAGGATATGAGCCCACT 838
 DB 19512 GAGCAGAGATTAGAGAGCAGAGGATTAGAGAGCAGAGGATTAGAGAGCAG 19453
 QY 839 TACCCCTGATATGAGCTCCATCTGCTGGAATATCAG 877
 DB 19452 CAGAGGTGAGAGCAGAGCAGAGCAGAGGAAAGCAG 19414

RESULT 8
 US-08-841-349-15/c
 Sequence 15, Application US/08841349B
 Patent No. 5955594
 GENERAL INFORMATION:
 APPLICANT: MISHRA, LOPA
 TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
 FILE REFERENCE: X/PO470050
 CURRENT APPLICATION NUMBER: US/08/841,349B
 CURRENT FILING DATE: 1997-04-30
 NUMBER OF SEQ ID NOS: 18
 SOFTWARE: Patent Ver. 2.0
 SEQ ID NO 15
 LENGTH: 981
 TYPE: DNA
 ORGANISM: Mus musculus
 FEATURE:
 OTHER INFORMATION: For all n's in this sequence, n=(a or g or c or t)
 US-08-841-349-15

Query Match 6.5%; Score 58.4; DB 2; Length 981;
 Best Local Similarity 49.8%; Pred. No. 8.3e-08;
 Matches 146; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 594 AGGATATGAGAGCCCAACAGGGGATATGAGAGCCCACTTATGAGATATGAGAGCCCGCC 653
 DB 944 AGCCATATGAGCAGCTCTCCAGGCGCATGATGACGCTCTCCAGGCGCATGATGACGCTCTC 885

QY 654 TGTGGATATGAGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCC 713
DB 884 TCCAGGCCATGTAGCAGCTCTCCAGGCCATGTAGCAGCTCTCCAGGCCATGTAGCAG 825
QY 714 ACCTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCC 773
DB 824 CTCTCAGAGCCATGTAGCAGCTCTCCAGGCCATGTAGCAGCTCTCCAGGCCATGTAGCAG 765
QY 774 CCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCC 833
DB 764 CAGATCTCAGAGCCATGTAGCAGCTCTCCAGGCCATGTAGCAGCTCTCCAGGCCATGTAGCAG 705
QY 834 AGCCCTACCCCTGCATATGAGTCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCC 886
DB 704 GGGCAGCTCTCAGAGCCATGTAGCAGCTCTCCAGGCCATGTAGCAGCTCTC 652

RESULT 9
US-08-917-320-18
Sequence 18, Application US/08917320
Patent No. 5824508
GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: No. 5824508 Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSES: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/917,320
FILING DATE: 25-AUG-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Luann Cseer
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
TELEX: 380816 CooleyPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURES:
NAME/KEY: CDS
LOCATION: 1014..3734
US-08-917-320-18

Query Match 6.1%; Score 54.8; DB 1; Length 3833;
Best Local Similarity 48.7%; Pred. No. 2e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
QY 516 AGCATATTCATTTGATCTATGAGCCCAACACAGAGATATACAGTCCACAGAGGGA 575
DB 2532 ACCAGCTCACTCAACAGTACTACCAACCCCAAAAGCAGCAGCCCAACCCAGAGA 2591

QY 576 ATATGAACTCCACAGAGATATGAGCCCAACAGGGGATATGAGCCCACTAT 635
DB 2532 GTGACTATCCCAACCCCAAAAGCAGCCCAACAGGATATACCAACCCCA 2651
QY 636 GGGATATGAGCCCGCTGTGGATATGAGTCCCACTGGGGATATGAGTCCCA 695
DB 2652 AATGCCACAGCCCACTTGGGAAAACAAAGTCTTACAGAGTACCTACCCCA 2711
QY 696 TGGGGATATGAGTCCCACTTGGGGATATGAGCCCACTTGGGGATATGAGTCC 755
DB 2712 CCAAAATGCCACAGCCCACTTGGGAAAACAAAGTCTTACAGAGTACCTACCCCA 2771
QY 756 ACCTGGGGATATGAGTCCCACTTGGGGATATGAGCCCACTTGGGGATATGAGTCC 815
DB 2772 ACCCAAAATGCCACAGCCCACTTGGGAAAACAAAGTCTTACAGAGTACCTACCCCA 2831
QY 816 CCCACC 821
DB 2832 CCAACC 2837

RESULT 10
PCT-US95-04611A-18
Sequence 18, Application PC/TUS9504611A
GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: Non Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSES: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04611A
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Luann Cseer
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
TELEX: 380816 CooleyPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURES:
NAME/KEY: CDS
LOCATION: 1014..3734
PCT-US95-04611A-18

Query Match 6.1%; Score 54.8; DB 5; Length 3833;
Best Local Similarity 48.7%; Pred. No. 2e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
QY 516 AGCATATTCATTTGATCTATGAGCCCAACACAGAGATATACAGTCCACAGAGGGA 575

Db 2532 ACCAGCTCCACCTGACGACTACCCCAATGCCAAGCCCAAGCCCAAGCA 2591
Qy 576 ATATGGAATCTCCACGAGAAATATGAGCCCAACAGGGGGATATGAGCCCACTAT 635
Db 2592 GTGACTACCCCAACCCCAATGACCAAGCCCAAGCAAGTACTACCCCAAGCCCA 2651
Qy 636 GGAATATGAGAGCCCGCTGTGGGATATGAGTCCCACTGGGGGATATGAGTCCACC 695
Db 2652 AATGCAACCAAGCCCACTGTGGGAAAAAACAAGCTCACTGACAGTACTACCCCAACC 2711
Qy 696 TGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGCCCAAGTATGAGTCCACC 755
Db 2712 CCAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGACAGTACTACCCCA 2771
Qy 756 ACCTGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGTATGAGTCCACC 815
Db 2772 ACCCAAAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGACAGTACTACC 2831
Qy 816 CCCACC 821 0
Db 2832 CCAACC 2837

RESULT 11

US-08-783-774-1

Sequence 1, Application US/08783774

Patent No. 6054130

GENERAL INFORMATION:

APPLICANT: Spaete, Richard

APPLICANT: Jackman, Winthrop

TITLE OF INVENTION: NON-SPLICING VARIANTS OF

TITLE OF INVENTION: GP350/220

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSER: Penile & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036/2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/783,774

FILING DATE: 15-JAN-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7682-037

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090

TELEFAX: 212-869-8864

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5931 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 1014...3734

OTHER INFORMATION:

US-08-783-774-1

Query Match 6.1%; Score 54.8; DB 3; Length 5931;
Best Local Similarity 48.7%; Pred. No. 2.4e-06;

Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
Qy 516 AGCATATTCMAATGATCTATGAGCCCAACAGAGATATACGTCAACAGAGGA 575
Db 2532 ACCAGCTCCACCTGACGACTACCCCAATGCCAAGCCCAAGCCCAAGCA 2591
Qy 576 ATATGGAATCTCCACGAGAAATATGAGCCCAACAGGGGGATATGAGCCCACTAT 635
Db 2592 GTGACTACCCCAACCCCAATGACCAAGCCCAAGCAAGTACTACCCCAAGCCCA 2651
Qy 636 GGAATATGAGAGCCCGCTGTGGGATATGAGTCCCACTGGGGGATATGAGTCCACC 695
Db 2652 AATGCAACCAAGCCCACTGTGGGAAAAAACAAGCTCACTGACAGTACTACCCCAACC 2711
Qy 696 TGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGCCCAAGTATGAGTCCACC 755
Db 2712 CCAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGACAGTACTACCCCA 2771
Qy 756 ACCTGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGTATGAGTCCACC 815
Db 2772 ACCCAAAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGACAGTACTACC 2831
Qy 816 CCCACC 821
Db 2832 CCAACC 2837

RESULT 12

US-09-556-706B-1

Sequence 1, Application US/09556706B

Patent No. 6458364

GENERAL INFORMATION:

APPLICANT: Spaete, Richard

APPLICANT: Jackman, Winthrop

TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220

FILE REFERENCE: 7682-050-999

CURRENT APPLICATION NUMBER: US/09/556,706B

CURRENT FILING DATE: 2000-04-24

PRIOR FILING DATE: 1997-01-15

PRIOR APPLICATION NUMBER: 08/783,774

PRIOR FILING DATE: 1994-04-18

NUMBER OF SEQ ID NOS: 19

SOFTWARE: PatentIn version 3.0

SEQ ID NO 1

LENGTH: 5931

TYPE: DNA

ORGANISM: Virus

FEATURE:

OTHER INFORMATION: GP350/220

US-09-556-706B-1

Query Match 6.1%; Score 54.8; DB 4; Length 5931;
Best Local Similarity 48.7%; Pred. No. 2.4e-06;
Matches 149; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
Qy 516 AGCATATTCMAATGATCTATGAGCCCAACAGAGATATACGTCAACAGAGGA 575
Db 2532 ACCAGCTCCACCTGACGACTACCCCAATGCCAAGCCCAAGCCCAAGCA 2591
Qy 576 ATATGGAATCTCCACGAGAAATATGAGCCCAACAGGGGGATATGAGCCCACTAT 635
Db 2592 GTGACTACCCCAACCCCAATGACCAAGCCCAAGCAAGTACTACCCCAAGCCCA 2651
Qy 636 GGAATATGAGAGCCCGCTGTGGGATATGAGTCCCACTGGGGGATATGAGTCCACC 695
Db 2652 AATGCAACCAAGCCCACTGTGGGAAAAAACAAGCTCACTGACAGTACTACCCCAACC 2711
Qy 696 TGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGTATGAGTCCACC 755
Db 2712 CCAATGCAACCAAGCCCACTGTGGGAAAAAACAAGCCCACTGACAGTACTACCCCA 2771
Qy 756 ACCTGGGGGATATGAGTCCCACTGGGGGATATGAGCCCAAGTATGAGTCCACC 815

Db 2772 ACCCAATGACACACCCCTTGGAAAAACAAGCCCACTGAGAGTACTACC 2831
Qy 816 CCCACC 821
Db 2832 CCAACC 2837

RESULT 13

US-08-182-175A-56
Sequence 56, Application US/08182175A
Patent No. 5559223
GENERAL INFORMATION:
APPLICANT: Saverio Carl Falco
APPLICANT: Sharon J. Keeler
APPLICANT: Janet A. Rice
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F
NUMBER OF SEQUENCES: 113
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. du Pont de Nemours and Company
STREET: 1007 Market Street
CITY: Wilmington
STATE: Delaware
COUNTRY: USA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh System, 6.0
SOFTWARE: Microsoft Word, 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/182,175A
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/743,006
FILING DATE: 9 August 1991
ATTORNEY/AGENT INFORMATION:
NAME: Linda Axamethy Floyd
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: BB-1031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (302) 992-4929
TELEFAX: (302) 892-7949
TELEX: 835420
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: B. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "sep"
OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-182-175A-56

Query Match 5.9%; Score 53.4; DB 1; Length 243;
Best Local Similarity 53.0%; Pred. No. 1.4e-06;
Matches 114; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 599 ATGGAGCCCAACGAGGGGATATGAGCCCACTATGAGTATGAGCCCGCGCTGTG 658
Db 2 ATGGAGGAGAGGCTGAAGCGATGAGAGGAGGAGGCTGAAGCGATGAGAGGAGCTGAAG 61

Qy 659 GATATGAGTCCACCTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 718
Db 62 GCATGAGAGGAGGCTGAAGCGATGAGAGGAGGAGGCTGAAGCGATGAGAGGAGCTG 121
Qy 719 GGGATATGAGAGCCCACTGAGGGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTG 778
Db 122 AGCGATGAGAGGAGGAGGCTGAAGCGATGAGAGGAGGAGGCTGAAGCGATGAGAGGAG 181
Qy 779 CTGAGGATATGAGAGCCCACTGAGGGGATATGAG 813
Db 182 CTCAAAAAGATGAGAGGAGGAGGCTTAATGATGAG 216

RESULT 14

US-08-474-633A-74
Sequence 74, Application US/08474633A
Patent No. 5773691
GENERAL INFORMATION:
APPLICANT: E. I. DU PONT DE NEMOURS AND
APPLICANT: COMPANY
TITLE OF INVENTION: CHIMERIC GENES AND
TITLE OF INVENTION: METHODS FOR INCREASING
TITLE OF INVENTION: INCREASING THE LYSINE
TITLE OF INVENTION: AND THRONINE CONTENT
TITLE OF INVENTION: OF THE SEEDS OF PLANTS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS
ADDRESSEE: AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U. S. A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD VERSION 2.0C
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,633A
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: BARBARA C. SIEGEL
REGISTRATION NUMBER: 30,684
REFERENCE/DOCKET NUMBER: BB-1037-C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-992-4931
TELEFAX: 302-773-0164
TELEX: 835420
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
STRAIN: B. coli
CELL TYPE: DHS alpha
IMMEDIATE SOURCE:
CLONE: 2-9
FEATURE:
NAME/KEY: CDS
LOCATION: 2..235
OTHER INFORMATION: /function= "synthetic
OTHER INFORMATION: storage protein"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "sep"
OTHER INFORMATION: /standard_name= "7.7.7.7.7.8.9.8.9.5"
US-08-474-633A-74

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: December 16, 2003, 10:43:10 / Search time 295.533 Seconds

(Without alignments)
10099.015 Million cell updates/sec

Title: US-09-864-291-4_COPY_36_933

Perfect score: 898
Sequence: 1 atggcagctgaccagcagcga.....tgctctcacagatctatga 898

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2201672 seqs, 166179599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262.6	29.2	467	11 US-09-764-891-1230	Sequence 1230, Ap
2	220	24.5	7099	11 US-09-764-891-7570	Sequence 7570, Ap
3	216.8	24.1	436	9 US-09-864-761-1561	Sequence 1561, A
4	195.4	21.8	321	11 US-09-764-891-1356	Sequence 1356, Ap
5	122	13.6	432	11 US-09-918-995-3356	Sequence 3356, A
6	114.4	12.7	409	10 US-09-960-352-6984	Sequence 6984, Ap
7	92.4	10.3	291	11 US-09-764-891-7571	Sequence 7571, Ap
8	86	9.6	471	9 US-09-864-761-15029	Sequence 15029, A
9	78.6	8.8	511	13 US-10-029-386-24981	Sequence 24981, A
10	77.8	8.7	593	13 US-10-027-632-290832	Sequence 290832, A
11	77.8	8.7	593	14 US-10-027-632-290832	Sequence 290832, A
12	73.6	8.2	420	10 US-09-974-300-3626	Sequence 3626, Ap
13	69.4	7.7	866	13 US-10-029-386-22839	Sequence 22839, A
14	68.8	7.7	591	9 US-09-864-761-19727	Sequence 19727, A
15	68.8	7.7	1959	9 US-09-864-761-2946	Sequence 2946, Ap

16	68.6	7.6	2016	10 US-09-938-842A-2004	Sequence 2004, Ap
17	63.6	7.1	1040	13 US-10-029-386-26033	Sequence 26033, A
18	61.6	6.9	541	13 US-10-029-386-11265	Sequence 11265, A
19	59.8	6.7	3469	13 US-10-029-804-1	Sequence 1, Appl1
20	59.4	6.6	541	13 US-10-029-386-12333	Sequence 12333, A
21	58.8	6.5	417	9 US-09-864-761-27093	Sequence 27093, A
22	58.8	6.5	3048	13 US-10-027-632-113439	Sequence 113439, A
23	58.8	6.5	3048	14 US-10-027-632-113439	Sequence 113439, A
24	57.2	6.4	511	13 US-10-029-386-24981	Sequence 24981, A
25	55.6	6.2	628	13 US-10-029-386-22859	Sequence 22859, A
26	54.8	6.1	2108	10 US-09-962-832-225	Sequence 225, App
27	54.6	6.1	417	9 US-09-864-761-27093	Sequence 27093, A
28	54.6	6.1	422	10 US-09-738-973-337	Sequence 337, App
29	54.6	6.1	422	10 US-09-854-133-337	Sequence 337, App
30	54.6	6.1	422	15 US-10-144-642A-337	Sequence 337, App
31	54.6	6.1	3048	13 US-10-027-632-113439	Sequence 113439, A
32	54.6	6.1	3048	14 US-10-027-632-113439	Sequence 113439, A
33	53.8	6.0	541	13 US-10-029-386-11265	Sequence 11265, A
34	53.4	5.9	243	15 US-10-023-066A-74	Sequence 74, Appl
35	53.2	5.9	3773	9 US-09-925-302-47	Sequence 47, Appl
36	53.2	5.9	6354	15 US-10-084-817-158	Sequence 158, App
37	52.6	5.9	354	9 US-09-864-761-18194	Sequence 18194, A
38	52.6	5.9	431	13 US-10-029-386-14222	Sequence 14222, A
39	52.6	5.9	597	13 US-10-029-386-517	Sequence 517, App
40	52.6	5.9	2030	10 US-09-880-107-2416	Sequence 2416, Ap
41	52.6	5.9	8923	10 US-09-764-847-1041	Sequence 1041, Ap
42	52.6	5.9	8923	15 US-10-092-154-1041	Sequence 1041, Ap
43	52.4	5.8	2746	13 US-10-371-723-1	Sequence 1, Appl1
44	52.2	5.8	689	13 US-10-027-632-148124	Sequence 148124, A
45	52.2	5.8	689	14 US-10-027-632-148124	Sequence 148124, A

ALIGNMENTS

RESULT 1
US-09-764-891-1230
Sequence 1230, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1230
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FRAGMENT:
NAME/KEY: SITE
LOCATION: (421)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1230

Query Match 29.2% Score 262.6; DB 11, Length 467;
Best Local Similarity 76.4% Pred. No. 2.1e-77;
Matches 365; Conservative 0; Mismatches 100; Indels 13; Gaps 3;
QY 351 CAGAGAGAGAGTGCATGCAATTCGCCAATGATGTAAGAGCTGCTCTGCTGC 410
DB 2 CAGAAATGAGAGTGCATGCAATTCGCCAATGATGTAAGAGCTGCTCTGCTGC 61
QY 411 CAGAGAAATTCACCTTGAAGTGAATTAATTCCTTGAACATTCAGACTGTAACAT 470
DB CCAGAGATTCACCTTGAAGTGAATTAATTCCTTGAACATTCAGACTGTAACAT 121
QY 471 TACTGTCACAGAGGCTGAGTGTCTTCAGAGACACTGTCAGCATATTCAT 530
DB 122 TACTGTCACAGAGGAAAT---ATGTGCACTCCACAGAGTCTTGTTCAG-----TTAT 169

Oy	53	GATCTATGAGCCCCCAACAGAGATATCACTCCAAACAGAGGAAATATGAACTCCACC	590
Db	170	TGCTCTATGAGCCCCCACTCGAGAGATATGAGAGCCCCCACTCCCGGATACGAGGCCCAACC	229
Oy	591	AGAGAGATATGAGCCCAACAGAGGAGATATGAGAGCCCCCACTATGAGATATGAGGCCCC	650
Db	230	TGCAGGATATGAGGCCCAACCCCTAGGAATATAGAGCCCGCTGTGGATTCAGAGCTTC	289
Oy	651	GCCGTGGAGTATGAGAGTCCCACTCGAGGAGATATGAGAGTCCCACTGGAGGATATGAGAGT	710
Db	290	ACCTGTGCATATGAGAGCCCCCACTCTTGAGATACGAGGCCCACTCGAGAGATATGAGAGC	349
Oy	711	CCCACTGGAGGATATGAGAGCCCCC-ACCTGGAGGATATGAGAGTCCCACTGGAGGATATG	769
Db	350	CCCACTCTGAGGATATGAGAGCCCAACTTGTGATATGAGAACCCCACTCTCGAGATATG	409
Oy	770	GTGCCCCCACTGGAGGATATGAGAGCCCCCACTCGAGAGATATGAGAGCCCCCACTGAG	827
Db	410	GAGCCCACTTATCGATATGAGAGCCCCCACTCGAGAGAAATGAGAGCCCGCTTTCGAG	467

```

RESULT 2
US-09-764-891-7570
: Sequence 7570, Application US/09764891
: Publication No. US20030077808A1
GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OP INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC006
: CURRENT APPLICATION NUMBER: US/09/764, 891
: PRIORITY FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OP SEQ ID NOS: 10231
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 7570
: LENGTH: 7099
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-891-7570

```

Query Match	24.5%	Score 220;	DB 11;	Length 7099;
Best Local Similarity	75.3%;	Pred. No. 2.3e-62;		
Matches 274;	Conservative 0;	Mismatches 90;	Indels 0;	Gaps 0;

OY	ATATCAATTTGGATCTATGAGCCCCCAACACAGATATTAAGTTCACCAACGAGGGGAATTA	578
Dd	ATTCACAGTATTTGTCTATGAGCCCCCACTGCAGAGATATGAGCCCCCACTCCCGGATTA	6777
OY	TGGAATCTCACACAGAGATATGAGCCCAACAGGGGAGATATGAGCCCCCACTTATGGG	638
Dd	CGAGGCCCACTGTGAGGATATGAGCCCAACCGTAAAGAAATGAAGCCCCGGCTGTGGG	6833
OY	ATTATGAGCCCCCGCTGTGGGATATGAGTCCCACTGGGGGATATGAGTTCACCTTGG	698
Dd	ATACAGAGCTTCACCTGTGCGGATATGAGCCCCCACTTGGGATATCGAGGCCCACTTGGC	6897
OY	GGGATATGAGAGTCCCACTGGGGGATATGAGCCCCCACTGGGGGATATGAGTTCACAC	758
Dd	AGGATATGAGGCCCACTCTAAGATATGAGCCCCCACTCTTGGATATGAGAACCCAC	6955
OY	TGGGGGATATGTGTCGCCCACTTGGGGGATATGAGCCCCCACTGCAGATATGAGAGCCC	818
Dd	TCTGGGATATGAGGCCCACTCTGGGATATGAGAGCCCCCACTGCAGGAATGAAGGCC	7017
OY	ACCACTGGAATGAGCCCTTACCCCTGCATATGAGGCTCCATCTGCTGGAATACAC	878
Dd	GCTTCGGGATATCAGAGCTTCACCTGTGGATATGAGAGCCAGGCTCAGAGATCTACAGC	7077
OY	TGCC	882
Dd	AGCC	7081

RESULT 3
US-09-864-761-31561
; Sequence 31561, Application US/09864761
; Patent No. US20020048763A1
END: 20020048763A1

```

1 GENERAL INFORMATION:
2 APPLICANT: Penn, Sharon G.
3 APPLICANT: Rank, David R.
4 APPLICANT: Hanzel, David K.
5 APPLICANT: Chen, Wensheng
6 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
7 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

```

FILE REFERENCE: A60MICA-X-1
CURRENT APPLICATION NUMBER: US/09/864,761

! CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04

; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-10-04

! PRIOR FILING DATE: 2000-09-21
! PRIOR APPLICATION NUMBER: PCT/US01/0066
! PRIOR FILING DATE: 2001-03-20

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/0066
PRIORITY DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/0066
PRIORITY DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/0066
PRIOR FILING DATE: 2001-01-30

; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/0066
 ; PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/0066
PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/0066
PRIOR FILING DATE: 2001-01-30

; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/0066
 ; PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/0066
PRIOR FILING DATE: 2001-01-30

; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/0067

;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-08-31

; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-05-20

; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-30

; PRIORITY FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117

```

; SOFTWARE: Anomax sequence listing Eng1
; SEQ ID NO 31561
      EVNCTM: 435

```

LENGTH: 436
TYPE: DNA
ORIGIN: HERR

ORGANISM: HOMO SAPIENS
 FEATURE:
 OTHER INFORMATION: MDD TO 700716 A

OTHER INFORMATION: EXPRESSED IN FETAL
OTHER INFORMATION: EXPRESSED IN PLACENTA

OTHER INFORMATION: EXPRESSED IN BONE MARROW

OTHER INFORMATION: SWISSBROT HIT: P1965

US-09-864-761-31561

Query Match	24.1%	Score 21
Best Local Similarity	75.6%	Prod No

Matches 269; Conservative 0; Mlma

527 TTGTGATCTATGGACCCCAACCAACGAG

527 TTGTGATCTATGACCCCCACCAACGATATACAGTCCAAACGAGGGAATATGCACTC 586

```

Db      1  TTAATGCTATGAGCCCACTGACAGATATGAGCCCACTCCGATACGAGCCC 60
Qy      587  CACGAGAAATATGAGCCCAACGAGGGGATATGAGCCCACTATGAGATATGAG 646
Db      61  CACTGAGATATGAGCCCAACGAGGAAATGAGGCGCGCTGAGATACAGAG 120
Qy      647  CCCGCGCTGAGATATGAGTCCCACTGAGGATATGAGTCCCACTGAGGATATG 706
Db      121  CTTCACTGTCGATATGAGCCCACTCTTGAATACGAGCCCACTGAGATATG 180
Qy      707  GAGTCCCACTGAGGATATGAGCCCACTGAGGATATGAGTCCCACTGAGGAT 766
Db      181  GAGCCCACTGAGATATGAGCCCACTCTTGAATATGAGCCCACTCTCGAT 240
Qy      767  ATGTGCCCCCACTGAGGATATGAGCCCACTGAGGATATGAGCCCACTGAG 826
Db      241  ATGAGGCCCACTCTGAGATATGAGCCCACTGAGGAAATGAGGCGCGCTGAG 300
Qy      827  GAAATGAGCCCTACCCCTGATATGAGGCTGATGCTGAAATACAGCTGCG 882
Db      301  GATACAGGCTCTGCTGATATGAGGAGGAGGCGCTGAGGAAATCTACAGAGCC 356

```

RESULT 4

```

US-09-764-891-1356
; Sequence 1356, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIORITY FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1356
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (321)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1356

```

Query Match 21.8%; Score 195.4; DB 11; Length 321;

Best Local Similarity 77.4%; Pred. No. 6.8e-55;

```

Matches 246; Conservative 2; Mismatches 69; Indels 1; Gaps 1;
Qy      29  GCGGTGAGGGGCGCTGATCCCTGCGGAAAGTGTCTTGAACAGATGAGAGTGG 88
Db      1  GCGCGCGGAGCGCTGATCCCTGAACGATGAAGTCTTGAACGCGGTCTCGAATGG 60
Qy      89  ACCTGCTCTCTACAGAAACAGATGATCTTATCTTATATGCAAAAGAGAA 148
Db      61  AGCTCTCTCTCCACAGAGATCAAGAGCTCAATGTCTTATGTTGAAAGACAGAA 120
Qy      149  CGTTGTTCTGACTTACATACCGGAGTGTCTTGTGACTTCACTTATGTAATGAC 208
Db      121  CATGTGTTCTCACTTACATACCGGAGTGTCTTGTGACTTCACTTATGTAATGAC 180
Qy      209  TGCCTTCTTATGATGCGGTTTGGCGGATGATGATGATGATGATGATGATGAT 268
Db      181  TGTGTCTTATATATGCAATTTATCTGATGATGATGATGATGATGATGATGAT 240
Qy      269  TTGCCCCCACTACATTAAGAGAACATTCAGAGAGCTCCAGGTGTGTGTGTGTG 327
Db      241  TTGCTGAAACTTATTAAGAGAACATTTACAGAGAGCTTCATATGAGGCTGTGAG 300

```

```

Qy      328  CAAGCTGTTTAAATTA 345
Db      301  CAAGCTACTTTAAATTA 318

```

RESULT 5

```

US-09-918-995-33536
; Sequence 33536, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIORITY FILING DATE: 2001-07-30
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 33536
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(432)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

```

Query Match 13.6%; Score 122; DB 11; Length 432;

Best Local Similarity 59.4%; Pred. No. 4e-30;

```

Matches 247; Conservative 0; Mismatches 160; Indels 9; Gaps 2;
Qy      1  ATGCAAGTGAACCAAGACCAACGAGGCGGTGAGGCGCTTATCCCTCTGAGAA 60
Db      18  ATGCGCGCTCAACAAATATCATCTGAGGCGGCGAGTATGTCATATAC---GAG 74
Qy      61  AGTGTCTTGAAGAGTGTGAGATGAGACCTTCTCTTACAGAAACAGTGAATCC 120
Db      75  AGCATCTTATAGCTCATATGATCACTGAGAACTCATATGATCAATGAAGAGGCCA 134
Qy      121  TATCTCTTAAATGAGCAAAAGAAAGAGTGTCTTCACTTACATACCGGAGTCTTC 180
Db      135  GAAGCCTTCAAGAGAACAAAGAGGACGTCTTACCTTACCCCTTACCGGATCTTT 194
Qy      181  GTGACTTCACTATGATGATGAGCCCAAGCTTTCTTATATGAGCGTGTGCTGATG 240
Db      195  CTGTC-----CAAGGCAAGAGATGCAATGCACTCTTCAATGATGCAATTTATCTCAG 248
Qy      241  AGTACTGACCATTAAGAACCAATTTTGGCCCCCACTTAAAGAACCATTCAG 300
Db      249  AAAGACTGTGATGATCAAGAGCGCGTATTTGTGCAATCAATCAAGAGGAAACAGTGAAG 308
Qy      301  GCAGCTCCAGTGTGAGTGTGAGAGCAAGCTGTTTAAATTTATCTTACAGAAAGAA 360
Db      309  GCGAAAGCGGAGAGTGTGAGAGGCTGCTTCTTCAAGATGATCTTCAAGCGGAGG 368
Qy      361  GTGCGCATGAATTTGCCCCCACTGATGATGATGATGATGATGATGATGATGAT 416
Db      369  GCGCGCATTAAGTTCGAGACAGCGGATGCTTCAAGTGTGATGATGATGATGATGAT 424

```

RESULT 6

```

US-09-960-352-6984
; Sequence 6984, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathiasagan, Nagapan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

```

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 6984
LENGTH: 409
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 30-LIB3058-024-01-K1-H9
US-09-960-352-6984

Query Match 12.7%; Score 114.4; DB 10; Length 409;
Best Local Similarity 59.4%; Pred. No. 1.4e-27;
Matches 234; Conservative 0; Mismatches 151; Indels 9; Gaps 2;

QY 1 ATGGAGTGAACGAGCCACACCGAGCGCTGCGGCGCCCTCATCCCTCTGGCGAA 60
DB 23 ATGGCTCATCAAGAACCACTCGAGGCGCGGAGTGATGTCACAAACAC--CGAG 79
QY 61 AGTGTCTTGAAGCACTGTGAAGATGTGACCTCTGCTCTTACAGAAACAGTGAATCC 120
DB 80 AGCATCTTGATGTCTTATGACCATGTAGAGCTTACGTTACAGATAGAGAAATGCCA 139
QY 121 TATCTTTAATGACAAAGAAAGAAAGTGTCTTCTCATCTTACCGGCTGCTTTC 180
DB 140 GAGGCTTCAAGGAGCAAGAAAGCAACCGCTTACCTTACCCCGTACCGGTACTTT 199
QY 181 GTGACTTCACTTATGATGATGATGATGATGATGATGATGATGATGATGATG 240
DB 200 CTGTC-----CAAGCGAGGAGATGATGATGATGATGATGATGATGATGATG 253
QY 241 AGTGTCTGACCATTTGAACAAACAAATTTTGGCCCACTTACATTAAGAAACATTCAG 300
DB 254 AAGGACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 313
QY 301 GCAAGCTGCAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 360
DB 314 GCCGAAAGCAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 373
QY 361 GGTGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 394
DB 374 GCGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 407

RESULT 7
US-09-764-891-7571
Sequence 7571, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7571
LENGTH: 291
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-7571

Query Match 10.3%; Score 92.4; DB 11; Length 291;
Best Local Similarity 76.0%; Pred. No. 2.9e-20;
Matches 114; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 171 GGTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 230
DB 9 GGTGATTTTCATTAATTCATGCTCCATGATGATGATGATGATGATGATGATG 68
QY 231 TGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 290

DB 69 TGATCTATGACGAACTTCACTGTGTAACACAGATATTTGTCGAACATTCATTAAGG 128
QY 291 AACCATCAGGACGCTCCAGGTGGTGGCTG 320
DB 129 AACTATTGAGGACGCTCCATATATGATG 158

RESULT 8
US-09-864-761-15029

Sequence 15029, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecomica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

PRIOR FILING DATE: 2001-05-23

PRIOR APPLICATION NUMBER: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annonax Sequence Listing Engine vers. 1.1

SEQ ID NO 15029

LENGTH: 471

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO 299716.4

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2

Query Match 9.6%; Score 86; DB 9; Length 471;

Best Local Similarity 77.6%; Pred. No. 5,4e-18;
Matches 104; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 627 CCCACCTATGAGATATGAGAGCCCGCTGTGGATATGAGTCCACCTGAGGAGATATG 666
DB 338 CCCAGTATATGCTATATGAGAGCCCGCTGTGGATATGAGTCCACCTGAGGAGATATG 397
QY 687 AGTCCCACTGAGGAGATATGAGTCCACCTGAGGAGATATGAGTCCACCTGAGGAGAT 746
DB 398 AGCCCACTGAGGAGATATGAGTCCACCTGAGGAGATATGAGTCCACCTGAGGAGAT 457
QY 747 TGAAGTCCCACTG 760
DB 458 CAGAGCTCACTG 471

RESULT 9

US-10-029-386-24981
Sequence 24981, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AROMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 24981
LENGTH: 511
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO 297205.1
US-10-029-386-24981
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48

Query Match 8.8%; Score 78.6; DB 13; Length 511;

Best Local Similarity 49.9%; Pred. No. 1.8e-15;
Matches 198; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 485 CTCGAGTGTCTCTCAACAGACCTGTGTCAGAGATATCCAAATGTGATCTATGACCC 544
DB 106 CTCGAGGCCCCCTCTCTCAACAGAGTCCCTCTCTCAACGAGGCTCCCTCTC 165
QY 545 CACCAACAGATATACAGTCCAAACAGAGGATATGAGTCCCAAGAAAGATATGAG 604
DB 166 CACCTGAGGCTCCCTCTCTCAACAGAGTCCCTCTCTCAACGAGGCTCCCTC 225
QY 605 CCAACAGAGGAGATATGAGGCTCCCTATGAGATATGAGGCTCCCTCTCTGATATG 664
DB 226 CTCGAGTGTCT 285
QY 665 GAGTCCCACTGAGGAGATATGAGTCCCACTGAGGAGATATGAGTCCCACTGAGGAT 724
DB 286 CTCGAGTGTCT 345
QY 725 ATGAGAGCCCACTGAGGAGATATGAGTCCCACTGAGGAGATATGAGTCCCACTGAGG 784
DB 346 CCTC 405
QY 785 GATATGAGGAGGAGATATGAGTCCCACTGAGGAGATATGAGTCCCACTGAGGAG 844
DB 406 GCTTC 465
QY 845 CTCGATATGAGTC 881
DB 466 CAGGCTC 502

RESULT 10
US-10-027-632-290832
Sequence 290832, Application US/10027632
Publication No. US20030204075A3
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 290832
LENGTH: 593
TYPE: DNA
ORGANISM: Human
US-10-027-632-290832

Query Match 8.7%; Score 77.8; DB 13; Length 593;

Best Local Similarity 54.9%; Pred. No. 3.6e-15;
Matches 151; Conservative 1; Mismatches 123; Indels 0; Gaps 0;

QY 142 AAGGAGAGTGTCTCTCACTCATACCGAGGTCTTGTGACTTCACTTATGATCAT 201
DB 45 AAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAGGAGAG 104
QY 202 GAGCCGAGCTTCTTTATGAGTCCGCTTGTGAGTATGAGTATGAGTATGAGTATG 261
DB 105 GATGCAAGAGGCT 164
QY 262 CCAATTTTGTCCCACTATATTAAGAACATTCAGGAGCTCCAGGTGAGCTG 321
DB 165 CCGTGTGAGCAAGTATGATCAAGGAGAGGAGAGGAGAGGAGAGGAGAGGAG 224
QY 322 GAGGAGAGGAGTATTAAGTATCTTCAAGAAAGAGTGCATCAATTTGCCAA 381
DB 225 GAGGAGTGTCT 284
QY 382 CTGATGTAAAGCT 416
DB 285 CAGGTCTCAAGTATCAATCTCAAGGCTCAAGG 319

RESULT 11

US-10-027-632-290832
Sequence 290832, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483

```

? PRIOR FILING DATE: 2000-03-29
? PRIOR APPLICATION NUMBER: US 60/185,218
? PRIOR FILING DATE: 2000-02-24
? PRIOR APPLICATION NUMBER: US 60/167,363
? PRIOR FILING DATE: 1999-11-23
? PRIOR APPLICATION NUMBER: US 60/156,358
? PRIOR FILING DATE: 1999-09-28
? PRIOR APPLICATION NUMBER: US 60/146,002
? PRIOR FILING DATE: 1999-08-09
? NUMBER OF SEQ ID NOS: 325720
? SOFTWARE: PasteSeq for Windows Version 4.0
? SEQ ID NO 290832
?
? LENGTH: 593
? TYPE: DNA
? ORGANISM: Human
?
US-10-027-632-290832

```

Query Match	8.7%;	Score 77.8;	DB 14;	Length 593;
Best Local Similarity	54.9%;	Pred. No. 3.6e-15;		
Matches 151; Conservative	1;	Mismatches 123;	Indels 0;	Gaps 0;

Qy	142	AAAGAAAGCTGTTTCTGACTTCATACCGGGGAGCTTCGAGACTTCAATTAGTCAAT	201
Db	45	AAAGGAGCCAAAGAAAGGACCATCTACCTTACCCCTTTCATCTTCTCTCGAAGGCCAAG	104
Qy	202	GACCCCATGCTTCTTTTATGATGCCGTTGGCTGATAGATGACTGCACATTGAACAA	261
Db	105	GATGCCACGAGCTCTTCTGTAAATGTCATTTTATCTTGAAGGCTATAGATCAAGCAG	164
Qy	262	CCAAATTTTGGCCCCCACTACATTTAAAGAAACCATTCAGAGAGCTCAAGGTGGGTGG	321
Db	165	CCGTGTATTGACAAACATGCATCAAGGGAAACATGAACACCGAAGCAGGGGTGGCTGG	224
Qy	322	GAAGACAAGCTGTTTAAAGTATTCCTCAGAAAGAGGTGCCATGCAATTTGCCCAA	381
Db	225	GAAGGCTGTGCTTCGCGAAGTCCGCTTCAAGGCAAGGGGGGGGCCATGTAATCTGACAG	284
Qy	382	CTGATGTAAAGCTGCTCTGTGCTGCGCAAGG	416
Db	285	CAGGTGCTCAGGTGACATCTCAAGGTCCAAAGGG	319

```

RESULT 12
US-09-974-300-3626
: Sequence 3626, Application US/09974300
: Patent No. US20020146721A1
: GENERAL INFORMATION:
: APPLICANT: Betra, Randy M.
: TITLE OF INVENTION: Methods For Monitoring Multiple Genes
: FILE REFERENCE: 10085, 500-US
: CURRENT APPLICATION NUMBER: US/09/974,300
: PRIOR FILING DATE: 2001-10-05
: PRIOR APPLICATION NUMBER: 09/680,598
: PRIOR FILING DATE: 2000-10-06
: PRIOR APPLICATION NUMBER: 60/279,526
: NUMBER OF SEQ ID NOS: 8481
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3626
: LENGTH: 420
: TYPE: DNA
: ORGANISM: Bacillus licheniformis
US-09-974-300-3626

```

Query Match	8.2%	Score 73.6;	DB 10;	Length 420;
Best Local Similarity	50.3%;	Pred. No. 7.5e-14;		
Matches 181; Conservative	0;	Mismatches 179;	Indels 0;	Gaps 0

Oy 522 TCCAA TTGTGATCTATGAACCCACCACAAGATTATACAGTCCAACGAGGGAAATTATCG 58

Db 29 TCGAAGCGGCTGATTATGCACGTTCGAAGCGGCTGATTATGCACGTTCGAAGCGGCTGATTATCG 88

Oy	582	AACTCCACGAAAGATATGAGCCCAACGAGGGGATATGAGACCCCACTATGGGATA	6411
Db	89	CATCGAAGGGCTGATATGAGCACTGAAAGGGCTGATATGAGCACTGAAAGCGCTGATA	148
Oy	642	TGAGACCCCGCTGTGGATATGAGTCCCACTGAGGGGATATGAGTCCCACTGAGGGG	7011
Db	149	TGGCAGTCCAAAGCGCTGATATGAGCACTGAAAGCGCTGATATGAGCACTGAAAGCGCTG	208
Oy	702	ATATGAGATCCCACTGAGGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTGAG	7611
Db	209	ATATGAGATCCCAAGGGCTGATATGAGCACTGAAAGCGCTGATATGAGCTCTGATCCGG	268
Oy	762	GGGATATGAGTCCCACTGAGGGGATATGAGCCCACTGAGGGGATATGAGTCCCACTG	8211
Db	269	CTGATATGAGCACTGAAAGCGCTGATATGAGCACTGAAAGCGCTGATATGAGCACTGAGC	328
Oy	822	AGCTGAAATGAAAGCCCTTACCCCTGCAATATGAAAGCTCATCTCTGGAAATATACAGCTGC	8811
Db	329	CCGTTATATGACATCTCAATGCAAGGATATTCGGGTCGAAACCGGATGACATTCGACTGC	388

```

RESULT 13
US-10-029-386-22839/c
/ Sequence 22839, Application US/10029386
/ Publication No. US20030194704A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
/ FILE REFERENCE: AROMICA-X-2
/ CURRENT APPLICATION NUMBER: US/10/029,386
/ CURRENT FILING DATE: 2001-12-20
/ NUMBER OF SEQ ID NOS: 34288
/ SOFTWARE: Annuomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 22839
/ LENGTH: 866
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURES:
/ OTHER INFORMATION: MAP TO AC005630.1
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
/ OTHER INFORMATION: SWISSPROT HIT: Q08379, EVALU 4.00e-10
/ OTHER INFORMATION: NT HIT: g116161055, EVALU 4.00e-95
/ OTHER INFORMATION: EST_HUMAN HIT: BF828725.1, EVALU 0.00e+00
US-10-029-386-22839

```

Query Match	7.74;	Score 69.4;	DB 13;	Length 866;
Best Local Similarity	49.94;	Pred. No. 3e-12;		
Matches 175, Conservative	0;	Mismatches 176;	Indels 0;	Gaps 0

QY 536 ATGACCCCCACACAGGATATATAGTCAACAGGGAAATATGAACTCCACACAG 539
Db 509 ACGTAGCTCTCTCTCTGTTCACAGAGCTCTCTCTGTTCAGTAGACCTCTCTCTG 455
QY 596 GATATGAGATCCCAACAGGGGATATAGAGCCCCACTATGGGATATAGAACCCCGCTG 655
Db 449 TTCAACAGGCTCTCTCTCTGTTCAGTAGAGCTCTCTCTGTTCACAGAGCTCTCTC 397
QY 656 TGGATATAGAGTCCCACTGGGGATATAGAGTCCCACTGGGGATATAGAGTCCAC 715
Db 389 CTGTTCATATAGGCTCTCTCTCTGTTCAGTAGAGCTCTCTCTTTCACATAGCTCTC 330
QY 716 CTGGGGATATAGAGCCCCACTGGGGATATAGAGTCCCACTGGGGATATAGAGTCC 775
Db 329 CTCTGTTCATATAGGCTCTCTCTCTGTTCAGTAGAGCTCTCTCTTTCACAGAGCT 277
QY 776 CACTTGGGGATATAGAGCCCCACTTGCAGATATAGAGCCCCCACTTGAAATAGAG 835

Db 269 CTCCTCCTGTCACGTAGCTCTCTCTGTTACAGAGCTCTCTCTGTCATGATAG 210
 QY 836 CCCATCCCCGATATGAAAGTCACTGCTGAAATAGAGCTGCTCTC 886
 Db 209 CCTCTCCTGTCATGATAGCTCTCTCTGTTACAGAGCTCTCTCTC 159

RESULT 14

US-09-864-761-19727
 ; Sequence 19727, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aeonica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263,6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/234,687
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 19727
 ; LENGTH: 591
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURES:
 ; OTHER INFORMATION: MAP TO AC002041.1
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.5
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.9
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.2
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.2

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
 ; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 4.3
 ; OTHER INFORMATION: SWISSPROT HIT: P03200, EVALU8 3.00e-16
 US-09-864-761-19727

Query Match 7.7%; Score 68.8; DB 9; Length 591;
 Best Local Similarity 51.6%; Pred. No. 3.8e-12;
 Matches 157; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 567 ACCAGGGGAATATGAACTCCACCAAGATATGAGCCCAACAGGGGAATATGAGC 626
 Db 170 ACCGAGGGATATGATGATGATGATGATGATGATGATGATGATGATGATG 229
 QY 627 CCCAGCTATGAGATATGAGACCCCGCTGAGGATATGAGTCCACCTGGGGATATG 686
 Db 230 GTACCTGTGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 289
 QY 687 AGTCCCACTGGGGGATATGAGTCCCACTGGGGGATATGAGCCCACTGGGGGATA 746
 Db 290 GAGGTACCTGGGGATATGATGATGATGATGATGATGATGATGATGATGATGATG 349
 QY 747 TGAATCCACCTGGGGATATGATGATGATGATGATGATGATGATGATGATGATG 806
 Db 350 TGTAGAGTACGATGAGGATATGATGATGATGATGATGATGATGATGATGATGATG 409
 QY 807 ATATGAGCCCACTGAGCTGGAATATGAGCCCTAACCCCTGATATGAGCTCATCTGC 866
 Db 410 ATGTGAGGTGACCTGGGGATGATGATGATGATGATGATGATGATGATGATGATG 469
 QY 867 TGA 870
 Db 470 GGGA 473

RESULT 15

US-09-864-761-2946
 ; Sequence 2946, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aeonica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263,6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 16, 2003, 09:21:45 ; Search time 1848.09 seconds
(without alignments)
11809.735 Million cell updates/sec

Title: US-09-864-291-4_COPY_36_933

Sequence: 1 atgcgcagtgaccagagccca.....tgcctctcacagatcctatga 898

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estr1:*
2: em_estr2:*
3: em_estr3:*
4: em_estr4:*
5: em_estr5:*
6: em_estr6:*
7: em_estr7:*
8: em_estr8:*
9: gb_estr1:*
10: gb_estr2:*
11: gb_estr3:*
12: gb_estr4:*
13: gb_estr5:*
14: gb_estr6:*
15: em_estr7:*
16: em_estr8:*
17: em_estr9:*
18: em_estr10:*
19: em_estr11:*
20: em_estr12:*
21: em_estr13:*
22: em_estr14:*
23: em_estr15:*
24: em_estr16:*
25: em_estr17:*
26: em_estr18:*
27: em_estr19:*
28: gb_estr20:*
29: gb_estr21:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	505.8	56.3	1079	12	BM564167 AGENCOURT
2	414.4	46.1	1414	11	AKO15863 Mus muscu
3	340.8	38.0	663	10	BG699398 602679081
4	340	37.9	775	10	BG722815 602695278

Result No.	Score	Query Match	Length	DB ID	Description
5	333.8	37.2	717	14	BY715414
6	206	22.9	701	10	BG701881
7	189.8	21.1	360	13	BY098622
8	183.6	20.4	932	13	BU107273
9	179.4	20.0	560	13	BU103740
10	179.4	20.0	700	13	BU284977
11	179.4	20.0	723	13	AJ452591
12	177.8	19.8	568	12	BU390687
13	177.8	19.8	636	12	BM490729
14	167.4	18.6	772	9	AJ454950
15	166.8	18.6	706	13	BU433499
16	153.6	17.1	816	13	BU916890
17	153.6	17.1	931	13	BU914288
18	152.2	16.9	613	13	AL449211
19	152.2	16.9	647	9	AL449651
20	152.2	16.9	655	9	AL498723
21	152.2	16.9	661	9	AL497245
22	152.2	16.9	663	9	AL481810
23	152.2	16.9	663	9	AL496682
24	152.2	16.9	460	10	BS680329
25	152.2	16.9	509	10	BS15416
26	152.2	16.9	594	12	BM190972
27	152.2	16.9	607	12	BM190975
28	152.2	16.9	868	14	CD255258
29	151.6	16.9	460	10	BS13187
30	151.6	16.9	719	14	BS593075
31	150.6	16.8	611	9	AL446063
32	150.6	16.8	760	14	CS592279
33	150.4	16.7	726	14	CS591559
34	150.4	16.7	862	13	BO735296
35	150.4	16.7	892	14	CD301781
36	143.6	16.0	488	28	AQ997933
37	141.2	15.7	635	9	AL488954
38	137.2	15.3	814	13	BU961805
39	136.8	15.2	622	10	BR468886
40	135.6	15.1	899	10	BR470466
41	132.8	14.8	652	14	CA590725
42	131.8	14.7	631	13	BU805447
43	131.8	14.7	744	13	BU230211
44	131.2	14.6	635	13	BU807307
45	130	14.5	437	13	BY249862

ALIGNMENTS

RESULT 1
LOCUS BM564167 1079 bp mRNA linear EST 20-FEB-2002
DEFINITION AGENCOURT_6560149 NIH_MGC_119 Homo sapiens CDNA clone IMAGE:5742129
ACCESSION BM564167
VERSION BM564167.1 GI:18811738
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1079)
AUTHORS NIH-MGC http://mgi.nhl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM12759 row: m column: 10
High quality sequence stop: 726.

LOCUS BG693938 663 bp mRNA linear EST 07-MAY-2001
 DEFINITION 602679081P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5',
 mRNA sequence.
 ACCESSION BG693938
 VERSION BG693938.1 GI:13967653
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 663)
 NIH-MGC http://mgi.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished
 JOURNAL Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.llnl.gov
 Plate: LLM10703 row: a column: 21
 High quality sequence stop: 660.
 Location/Qualifiers
 1..663
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4811804"
 /label_type="h1pocampus"
 /lab_host="DH108"
 /clone_id="NIH_MGC_95"
 /note="Organ: Brain; Vector: pBluescript (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gscgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
 size-selected for average insert size 2.5 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 168 a 158 c 169 g 164 t 4 others
 ORIGIN
 Query Match 38.0%; Score 340.8; DB 10; Length 663;
 Best Local Similarity 75.0%; Pred. No. 9.8e-88;
 Matches 481; Conservative 0; Mismatches 145; Indels 15; Gaps 4;
 Oy 1 ATGGAGTGAACCAAGCAACCAAGAGCGCTGTGGGCGCTTCATCCCTCTGGCGAA 60
 Db 37 ATGGAGTGAACCAAGCAACCAAGAGCGCTGTGGGCGCTTCATCCCTCTGGCGAA 96
 Oy 61 ATGTCTTGAAGCAAGTGAAGATGTGACCTCTGCTTCAAGAAACAGATGAATCC 120
 Db 97 ATGTCTTGAAGCAAGTGAAGATGTGACCTCTGCTTCAAGAAACAGATGAATCC 156
 Oy 121 TATCTCTTAATAGCAAGAAAGAAAGAAAGTGTCTTCACTTCAATACCGGTGTCTTC 180
 Db 157 TATCTCTTAATAGTGAAGAAAGAAAGAAAGTGTCTTCACTTCAATACCGGTGTCTTC 215
 Oy 181 GTGACTTGAACCTTAAGTGAAGAAAGCAAGTGTCTTCACTTCAATACCGGTGTCTTC 240
 Db 216 ATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 275
 Oy 241 ATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 300
 Db 276 ATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 335
 Oy 301 GGAGCTCAAGT 360

Db 336 GCAGCTCATATGT 395
 Oy 361 GTGCTCATATGT 420
 Db 396 GTGCTCATATGT 455
 Oy 421 CCATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 480
 Db 456 CCATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 515
 Oy 481 GGGGCTGAGT 540
 Db 516 GGGGAT--ATGTGCTGT 563
 Oy 541 CCCCACACACAGAT 600
 Db 564 GCCCAGCTGT 621
 Oy 601 GGAGCCCAACCAAGGAGTATATATATATATATATATATATATATATATATATATAT 641
 Db 622 GGAGCCCAACCAAGGAGTATATATATATATATATATATATATATATATATATATAT 662
 RESULT 4
 Bg722815 775 bp mRNA linear EST 08-MAY-2001
 LOCUS Bg722815
 DEFINITION 602695278P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5',
 mRNA sequence.
 ACCESSION Bg722815
 VERSION Bg722815.1 GI:14002002
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 775)
 NIH-MGC http://mgi.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished
 JOURNAL Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNLN at:
 http://image.llnl.gov
 Plate: LLM10743 row: 1 column: 20
 High quality sequence stop: 651.
 Location/Qualifiers
 1..775
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4827427"
 /lab_host="DH108"
 /clone_id="NIH_MGC_97"
 /note="Organ: testis; Vector: pBluescript (modified
 pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gscgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
 size-selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 191 a 212 c 193 g 179 t
 ORIGIN
 Query Match 37.9%; Score 340; DB 10; Length 775;
 Best Local Similarity 75.5%; Pred. No. 1.8e-87;

Oy		AAATACATCTCCAGGGGCGTGCAGAGTGTCCTCCACAGACACTCTGTCCAGCATATTCCAAT	527
Oy		5468 AATTACATCTCCAGGGGCGTGCAGAGTGTCCTCCACAGACACTCTGTCCAGCATATTCCAAT	527
Db		516 AATTACATCT-GGAATTTGGAAATTTTGTCATCGTCCACAGATGCCCTTATCAG-----	564
Oy		528 TGTGATCTATGAGACCCCCACCAACAGATATATACAGTCCAACACAGGGAGTAATGGAATCC	587
Db		565 TTATGTCTATGAGAGCGCACACTCGGGAGTATGAGAGCCCCAAGCTCCGGA--TAGGAGAGCC	622
Oy		588 ACCAGAGAGATATGAGAGCCCAAACAGGGGGGATATGAGAGCCCACCTATGAGATATGAGAGC	647
Db		623 CACTGCAAGATATGAGAGCCAAACCCTGTGGAATATGAAAAGGCG-CGCTGTGGGATATCAGAGC	681
Oy		648 CCCGCGCTG 655	
Db		682 CTCACCTG 689	
RESULT 7			
LOCUS	BY098622	360 bp mRNA linear EST 07-DEC-2002	
DEFINITION	BY098622 RIKEN full-length enriched, pooled tissues, adult spleen, etc. Mus musculus cDNA clone K630123B09 5', mRNA sequence.		
ACCESSION	BY098622		
VERSION	BY098622.1 GI:26209239		
KEYWORDS	EST.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
REFERENCE	Bukacinska, Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 360) Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Oomoto, N., Satou, R., Suzuki, R., Yamanaoka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bulc, C., Hume, D.A., Quackelshuh, D., Schriml, L.M., Kanapin, A., Matsuda, H., Batzloff, S., Beisel, K.W., Blake, J.A., Bradd, D., Brunic, V., Choehia, C., Corbali, L.B., Cousins, S., Dalla, B., Dreagan, T.A., Fletcher, C.F., Forrest A., Frazer, K.S., Gasterland, T., Gatiboldi, M., Glasi, C., Godzik, A., Gough, J., Grimmond, S., Gutlinich, S., Hitokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasaki, Y., Kedziarski, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pereira, G., Pesole, G., Petrovsky, N., Pillai, R., Ponting, J.U., Qi, D., Ramachandran, S., Rayvasi, T., Reed, J.C., Reed, D.J., Reid, U., Ring B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou M., Shimada, M., Ventura, R., Takenaka, Y., Taylor, M.S., Teasdale R.D., Tomita, M., Urtado, R., Wagner, L., Wahlstedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wyshaw-Borja, A., Yanagisawa M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Zinnick, P., Hayatsu, N., Hirozane-Kobayashi, T., Komori, H., Nakamura M., Sekanuma, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Atzawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata K., Shingawa, A., Yasuniishi, A., Yoshino, M., Waterston, R., Lander R.S., Rogers, J., Birney, E. and Hayashizaki, Y.		
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs		
JOURNAL	Nature 420, 563-573 (2002)		
MEDLINE	Nature 420, 563-573 (2002)		
PUBMED	12466851		
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, url:http://genome.gsc.riken.go.jp/ Atzawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H., Miyazaki, A.,		

Murata, M., Nakamura, M., Nomura, K., Munazaki, R., Ohno, M., Sakai, X., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watabiki, A., Yamamoto, M., and Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site (<http://genome.gsfc.riken.go.jp>) for further details.

FEATURES
source

```

1..360
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="K630123B09"
/location="Riken full-length enriched, pooled tissues, adult spleen, etc."
/note="[dev_stage=adult, tissue_type=spleen, sex=male), (dev_stage=adult, tissue_type=Kidney, sex=male), (dev_stage=adult, tissue_type=testis, sex=male), (dev_stage=adult, tissue_type=thymus, sex=male), (dev_stage=adult, tissue_type=heart, sex=male), (dev_stage=adult, tissue_type=colon, sex=male), (dev_stage=adult, tissue_type=stomach, sex=male), (dev_stage=adult, tissue_type=liver, sex=male), (dev_stage=13 days embryo, tissue_type=whole body, sex=mix), (dev_stage=14 days embryo, tissue_type=whole body, sex=mix), (dev_stage=15 days embryo, tissue_type=whole body, sex=mix), (dev_stage=15 days pregnant, adult, tissue_type=amion, sex=female), (dev_stage=10 days neonate, tissue_type=brin, sex=mix), (dev_stage=10 days neonate, tissue_type=thymus, sex=mix), (dev_stage=10 days neonate, tissue_type=heart, sex=mix)]"
  
```

BASE COUNT 89 a 91 c 94 g 86 t

ORIGIN

```

Query Match      21.1%; Score 189.8; DB 13; Length 360;
Best Local Similarity 76.4%; Pred. No. 5.5e-44;
Matches 233; Conservative 0; Mismatches 72; Indels 0; Gaps 0
  
```

1 ATGGCAGTGAACCAAGGCCACACCGAGAGCCGTGTGGGGCCCTCATCCCTCTGGCGAA 60
 56 ATGGCAGTGAACCAAGGCCACATACAGTGAACCGCCGCTTGGGCGCCATCCCTCATGTGAA 115
 61 AGTGTCTGAAGCAGTGTGAAGAGTGTGACCTCTGCTCTCTACAGAAACCAAGTGAATCC 120
 116 AGTCTCTGAAGAAATGTTCAGAAAGTGAACCTCTCTCCCACTACACCGCAGAGCTCC 175
 121 TATCTCTTAATGCGACAAAGAAAGAAACGTGTGTTCACACTTACACCGGTGGCTTC 180
 176 AACCTCTTATGTGTGAACAAAGAGGGGGCTCTGTTCACACTTACACCGGTGAATTTT 235
 181 GTGACTTCACACTTAAGTCAATGACACCCAGCTTCTTTTATGATGCCGTTTGGACTGATG 240
 236 GTGACTTCGCGTGAACAAAGATCCCACTGTCTTTCTTTAAGATGCCATTTTCATCTAATG 295
 241 AGTACGTGCAACATTTGAACCAACCAATTTTGGCCCACTACATTAAGAAACCATTTGAG 300
 296 AATAACTGCACTGTGTGAACCAACATCTTTGGTGTGAACATCAATTTAAAGGACCAATTCAG 355

QY 301 GCAGC 305
 DB 356 GCAGC 360
 RESULT 8
 LOCUS BU107273 932 bp mRNA linear EST 25-NOV-2002
 DEFINITION 60311954P1 CSBQCHL12 Gallus gallus cDNA clone CHEST60k13 5', mRNA
 sequence.
 ACCESSION BU107273.1 GI:25309754
 VERSION BU107273.1
 KEYWORDS Gallus gallus (chicken)
 SOURCE Gallus gallus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 932)
 Boardman, P.B., Sanz-Bzquez, J., Overton, I.M., Burt, D.W., Bosch, B.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.,
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 JOURNAL MEDLINE 2235534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..932
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST60k13"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_1lb="CSBQCHL12"
 /note="Organ: heart; Vector: pBluescript II KS(+); Site 1:
 BcoRI; Site 2: NotI; Modification of Bluescript II KS(+)
 (Stratagene) vector to accommodate cDNA produced with the
 1-trimmed protocol (construction of uni-directionally
 cloned cDNA libraries from messenger RNA for improved 3'
 end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
 624). Cut pBluescript II KS(+) with NotI and BcoRI.
 ligate in double stranded adaptor containing BspGI and
 BamHI sites (5'gagcggctgcagcccgatccgaaaaaag)
 (5'aattctttctcgatccggcgctgcagc)"
 215 a 273 c 233 g 211 t
 BASE COUNT
 ORIGIN
 Query Match 20 4%; Score 183.6; DB 13; Length 932;
 Best Local Similarity 59.5%; Pred. No. 6,1e-42;
 Matches 350; Conservative 0; Mismatches 229; Indels 9; Gaps 2;
 QY 7 GTGAACGAGCCACGACGAGCGCTGCGGCGCTCATTCCTCTGGCGAAAGTGC 66
 DB 21 GGGCTCAATAGGACACCTCGAAGAGGCGGTGCTGCTCCCAATGCCAGAGCAT 80
 QY 67 TTGAAGCATGTGAGATGTGACCTCTGCTCTCAAGAAACGATGATCTATCTC 126
 DB 81 CTCAGCATGTAAAGATGTGAGCTCTCTCAAGTACATGACGGGCAAGCTTGAAGCC 140
 QY 127 TTAAATGGACAAAGAAAGAAAGTGTTCATCTCACTCAACGGGTGCTTGGAGCT 186
 DB 141 TTCAAGGACCAAGAAAGAAAGTGTATCTCACTCAAGGATGATCTTCTGTC- 199

QY 187 TCACATTAATGATGACCCGATGCTTTTATATATGATGCGCTGATGATGAC 246
 DB 200 -----AAGGCGAAGATCTTATGCTCTTTATATGATGCGCTTTATTTGGTGAAGG 254
 QY 247 TGACCATTTGAACCAATTTTGGCCCAACATTAAGGAACCATTCAGGAGCT 306
 DB 255 TGCTATTAAGAGCTCTTCTCTGCTTAATTAATCAAGGACGATTCAGGCTAG 314
 QY 307 CCAAGTGTGCTGAGGAAGACAGCTGTTTTTAAGTTATCTTCAGAAAGAGTGCC 366
 DB 315 GCAAGAGTGTGCTGAGGAAGGAGGAAGCTTTAACTGACTTTCAACAGGAGAGGCC 374
 QY 367 ATCAATTTGCCCATGATGTTAAAGCTGCTCTGCTGCTGCAGAGAAATTCATT 426
 DB 375 ATCAATTTGACAGCTGATGTTCAAGCTGCTCTGATGCTTCAATGAGATCTCTC 434
 QY 427 GAAAGTTAAATTAATGATGACACTTCAGACTGATCAATTAATTAATGATGAGGCT 486
 DB 435 CAGCCCCCTGCTATGATGATACACTGTAAGGATGATGACCTGCTCACTGCT 494
 QY 487 GCATGATGCTCTGACAGACAC--TTGTCCAGCATATCCATTTGATCTATGACCC 543
 DB 495 CCAAGAGGATGACCTGCTCCAGAGGATGACCTGCTCCAGAGGATGACCT 554
 QY 544 CCACCAAGATATGATGCTCAACAGGAGGATATGAACTCCACCA 591
 DB 555 GCTCCAGAGGATGACCTGCTCCGAGGAGGATGATGCTCTCTCTCA 602

RESULT 9
 LOCUS BU103740 560 bp mRNA linear EST 25-NOV-2002
 DEFINITION 603005979P1 CSBQCHL01 Gallus gallus cDNA clone CHEST32n20 5', mRNA
 sequence.
 ACCESSION BU103740.1 GI:25304305
 VERSION BU103740.1
 KEYWORDS Gallus gallus (chicken)
 SOURCE Gallus gallus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 560)
 Boardman, P.B., Sanz-Bzquez, J., Overton, I.M., Burt, D.W., Bosch, B.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.,
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 JOURNAL MEDLINE 2235534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..560
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST32n20"
 /issue_type="whole embryo"
 /dev_stage="20-21"
 /lab_host="DH10B"
 /clone_1lb="CSBQCHL01"
 /note="Organ: whole embryo; Vector: pBluescript II KS(+);
 Site 1: BcoRI; Site 2: NotI; Modification of pBluescript
 II KS(+) (Stratagene) vector to accommodate cDNA produced
 with the 1-trimmed protocol (construction of
 uni-directionally cloned cDNA libraries from messenger RNA

for improved 3' end DNA sequencing by Glenn Fu, et al.
 U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with
 NotI and EcoRI. Ligate in double stranded adaptor
 containing BspI and BamHI sites
 [5'ggcggcgctgcagcccgccgacccgaaaaag]
 [5'atctcttcttcggatccgggctgcagc]"

BASE COUNT

131 a 142 c 159 g 128 t

ORIGIN

Query Match 20.0%; Score 179.4; DB 13; Length 560;
 Best Local Similarity 66.1%; Pred. No. 7.5e-41;
 Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

7 GTGACCGAGGACCAACCGAGAGCCGTCGTGGGCGCCCTCATCCCTCTGGCGAAAGTGTG 66
 42 GCGCTCAATAGGACCACTCGAAGAGAGCGGTCTCTCTCCCAATGCCAGAGCAT 101
 67 TTGAGCAGTGTGAGATGTGACCTCTGCTTCTTACAGAAACGATGAACTCTATCTC 126
 102 CTCAGCAATGTAAAGATGTGAGCTCTCTTCAAGTACATGACGGGCAAGCTTGAAGCC 161
 127 TTTAATGGCAAAAGAAAGAAAGCTTTTCTGCTTCAACCGGGTGTCTTGACT 186
 162 TTCAGGACCAAGAAAGGAAATGTGATCTCAACCCATACAGAAATGATCTGTGTC- 220
 187 TCACACTAGTCAATGACCCCATGCTTTTATGATGACGATTTGGCTGATGATGAC 246
 221 -----AAGGCGAAGATCTCATGCTCTTTTATGATGACGATTTATTTGGTAAAGG 275
 247 TGACCATTTGAAACCAATTTTTCGCCCACTACATTAAGAAACATTCAAGCAGCT 306
 276 TGCTCATATGAGAGCCTCTTTCTGTCTATTAATACATCAAGAGCAGATTCAAGGCTAG 335
 307 CCAGGTGTGTGTGGGAAAGCAAGCTGTTTAAAGTATTCCTTCAAGAAAGAGGTGCC 366
 336 GCAGAGATGTGTGGGAAAGGAGGAGAAAGCTTTAACTGACTTTCACACAGCGAGAGCC 395
 367 ATGGAATTTGCCCACTGATGTAAAGTGCCTCTGCTGTCAGAGGAATTCGACT 425
 396 ATGAGTTTGACAGCTGATGTTCAGAGCTGCTTACGTCTTCAAGTGAATGTTCTCT 454

RESULT 10

BU284977

LOCUS 700 bp mRNA linear EST 27-NOV-2002

DEFINITION 603603379F1 CSBCHN54 Gallus gallus cDNA clone CHST58ln20 5', mRNA

ACCESSION BU284977

SEQUENCE

BU284977.1 GI:25724433

KEYWORDS

EST

SOURCE Gallus gallus (chicken)

ORGANISM

Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

1 (bases 1 to 700)

REFERENCE Bortman, P.R., Samz-Baquero, J., Overton, I.M., Burt, D.W., Bosch, R.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.,
 A Comprehensive Collection of Chicken CDNA

TITLE

JOURNAL

Curr. Biol. 12 (22), 1965-1969 (2002)

COMMENT

12445392

CONTACT: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1BD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

FEATURES

1..700

/organism="Gallus gallus"

/mol_type="mRNA"
 /strain="Compton Line 151"
 /db_xref="taxon:9031"
 /clone="CHST58ln20"
 /sex="Female"
 /tissue_type="not cerebrum or cerebellum"
 /dev_stage="adult"
 /lab_host="DH108"
 /clone_1lb="CSBCHN54"
 /note="Organ: brain; Vector: pBluescript II KS(+), Site_1:
 EcoRI Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunt-ended, ligated to NotI adaptors, digested with EcoRI
 , size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

BASE COUNT

162 a 196 c 178 g 164 t

ORIGIN

Query Match 20.0%; Score 179.4; DB 13; Length 700;
 Best Local Similarity 66.1%; Pred. No. 8.6e-41;
 Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

7 GTGACCGAGGACCAACCGAGAGCCGTCGTGGGCGCCCTCATCCCTCTGGCGAAAGTGTG 66
 6 GCGCTCAATAGGACCACTCGAAGAGAGCGGTCTCTCTCCCAATGCCAGAGCAT 65
 67 TTGAGCAGTGTGAGATGTGACCTCTGCTTCTTACAGAAACGATGAACTCTATCTC 126
 66 CTCAGCAATGTAAAGATGTGAGCTCTCTTCAAGTACATGACGGGCAAGCTTGAAGCC 125
 127 TTTAATGGCAAAAGAAAGAAAGCTTTTCTGCTTCAACCGGGTGTCTTGACT 186
 126 TTCAGGACCAAGAAAGGAAATGTGATCTCAACCCATACAGAAATGATCTGTGTC- 184
 187 TCACACTAGTCAATGACCCCATGCTTTTATGATGACGATTTGGCTGATGATGAC 246
 185 -----AAGGCGAAGATCTCATGCTCTTTTATGATGACGATTTATTTGGTAAAGG 239
 247 TGACCATTTGAAACCAATTTTTCGCCCACTACATTAAGAAACATTCAAGCAGCT 306
 240 TGCTCATATGAGAGCCTCTTTCTGTCTATTAATACATCAAGAGCAGATTCAAGGCTAG 299
 307 CCAGGTGTGTGTGGGAAAGCAAGCTGTTTAAAGTATTCCTTCAAGAAAGAGGTGCC 366
 300 GCAGAGATGTGTGGGAAAGGAGGAGAAAGCTTTAACTGACTTTCACACAGCGAGAGCC 359
 367 ATGGAATTTGCCCACTGATGTAAAGTGCCTCTGCTGTCAGAGGAATTCGACT 425
 360 ATGAGTTTGACAGCTGATGTTCAGAGCTGCTTACGTCTTCAAGTGAATGTTCTCT 418

RESULT 11

AJ452591

DEFINITION 723 bp mRNA linear EST 22-APR-2002

ACCESSION AJ452591

SEQUENCE

AJ452591.1 GI:20262687

KEYWORDS

EST

SOURCE Gallus gallus (chicken)

ORGANISM

Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

1 (bases 1 to 723)

REFERENCE Buerstedde, J.M.

TITLE Gallus gallus bursa lymphocyte EST
JOURNAL Unpublished
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinstr. 52, 20251 Hamburg, Germany
Email: URI: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.
Location/Qualifiers

FEATURES
source 1..723
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="31her1"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/clone_1ib="riken1"
/note="CB indred strain"

BASE COUNT 161 a 205 c 181 g 172 t 4 others
ORIGIN

Query Match 20.0%; Score 179.4; DB 9; Length 723;
Best Local Similarity 66.1%; Pred. No. 8.8e-41;
Matches 277; Conservative 0; Mismatches 136; Indels 6; Gaps 1;

QY 7 GTGAACGAGCCACGACCGAGCCGTCGTGGGCGCTTCATCCCTCTGGCGAAAGTGC 66
DB 15 GCGCTCAATAGGCACTCGAAGAGCGGTGTCGTCCCAATGCCAGAGCAATT 74
QY 67 TTGAAGCAGTGTGAGATGTGACCTCTGCTTCCACGAAACGATGGAATCTATCTC 126
DB 75 CTCAGGCAATGTAAAGATGTGAGCTCTCTCACTGATGACGAGCGGCAAGCTTAGGCC 134
QY 127 TTTAATGACAAAGAAAGAAAGCTGTTCTCACTTCATACCGGGTGTCTCTGTAAGT 186
DB 135 TTCAAAGGACCAAGAGGAATGCTGTATCTCAACCCATACAGATATCTTCGTGTC- 193
QY 187 TCACACTTAGTCATGATGACCCCATGCTTCTTTATGATGACCGCTTGGCTGTATGATGAC 246
DB 194 -----AAAGGCAAGATCTTATGCTGCTTTTATGATGACCGCTTATTTGATGAAAGG 248
QY 247 TGCACATTTGAACAACATTTTGGCCCACTACATTAAGAAACATTCAGGACGT 306
DB 249 TGCCTATTGAGCAGCTCTTCTTCTGTCTAATTCATCAAGAAAGCAAGATTCAGGCTGAG 308
QY 307 CCAGGTGTGCTGGGAGAGCAAGCTGTTTAAAGTATCCTTCAGAGAAAGAGTGC 366
DB 309 GAGAGAGTGTGCTGGGAGAGGAGGAGAACTTTAAAGTATTCATTCAGAGGAGAGCC 368
QY 367 ATCGAATTTGCCCACTGATGTAAAGCTGCTCTGCTGCGCAGAGAAATTCACCT 425
DB 369 ATCGAGTTTGAACAGCTGATGTTCAAGCTGCTCTAGTGTTCAGATGAGATTCCTCT 427

RESULT 12
BI390687 568 bp mRNA linear EST 06-AUG-2001
LOCUS BI390687
DEFINITION Pp2n.pk001.p22 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library Gallus gallus clone Pp2n.pk001.p22 5' similar
to gb|AA010951.1 (U79458) WW domain binding protein-2 [Homo sapiens],
mRNA sequence.
ACCESSION BI390687
VERSION BI390687.1 GI:15083969
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 568)
Porter T.B. and Cogburn L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
Library USDA/IFARs Animal Genome Project
JOURNAL Unpublished

COMMENT Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburnudel.edu, www.chickest.udel.edu.
Location/Qualifiers

FEATURES
source 1..568
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Commercial broiler chicken"
/db_xref="taxon:9031"
/clone="p2n.pk001.p22"
/sex="Male and Female"
/tissue_type="Pituitary gland/Hypothalamus/Pineal gland"
/dev_stage="Embryonic (d12,d14,d19); post-hatch (w1,w3,w5,w7,w9)"
/lab_host="B. coli BMDH108"
/clone_1ib="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"

BASE COUNT 128 a 146 c 159 g 130 t 5 others
ORIGIN

Query Match 19.8%; Score 177.8; DB 12; Length 568;
Best Local Similarity 65.9%; Pred. No. 2.2e-40;
Matches 276; Conservative 0; Mismatches 137; Indels 6; Gaps 1;

QY 7 GTGAACGAGCCACGACCGAGCCGTCGTGGGCGCTTCATCCCTCTGGCGAAAGTGC 66
DB 26 GCGCTCAATAGGCACTCGAAGAGCGGTGTCGTCCCAATGCCAGAGCAATT 85
QY 67 TTGAAGCAGTGTGAGATGTGACCTCTGCTTCCACGAAACGATGGAATCTATCTC 126
DB 86 CTCAGGCAATGTAAAGATGTGAGCTCTCTCACTGATGACGAGCGGCAAGCTTAGGCC 145
QY 127 TTTAATGACAAAGAAAGAAAGCTGTTCTCACTTCATACCGGGTGTCTCTGTAAGT 186
DB 146 TTCAAAGGACCAAGAGGAATGCTGTATCTCAACCCATACAGATATCTTCGTGTC- 204
QY 187 TCACACTTAGTCATGATGACCCCATGCTTCTTTATGATGACCGCTTGGCTGTATGATGAC 246
DB 205 -----AAAGGCAAGATCTTATGCTGCTTTTATGATGACCGCTTATTTGATGAAAGG 259
QY 247 TGCACATTTGAACAACATTTTGGCCCACTACATTAAGAAACATTCAGGACGT 306
DB 260 TGCCTATTGAGCAGCTCTTCTTCTGTCTAATTCATCAAGAAAGCAATTCAGGCTGAG 319
QY 307 CCAGGTGTGCTGGGAGAGCAAGCTGTTTAAAGTATCCTTCAGAGAAAGAGTGC 366
DB 320 GAGAGAGTGTGCTGGGAGAGGAGGAGAACTTTAAAGTATTCATTCAGAGGAGAGCC 379
QY 367 ATCGAATTTGCCCACTGATGTAAAGCTGCTCTGCTGCGCAGAGAAATTCACCT 425
DB 380 ATCGAGTTTGAACAGCTGATGTTCAAGCTGCTCTAGTGTTCAGATGAGATTCCTCT 438

RESULT 13
BM490729 636 bp mRNA linear EST 07-FEB-2002
LOCUS BM490729
DEFINITION Pp2n.pk004.e6 Normalized Chicken Pituitary/Hypothalamus/Pineal
Library (Pp2n) Gallus gallus cDNA clone Pp2n.pk004.e6 5' similar
to gb|AA010951.1 (U79458) WW domain binding protein-2 [Homo sapiens]
1. mRNA sequence.
ACCESSION BM490729
VERSION BM490729.1 GI:18611660
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

AUTHORS Boardman, P. B., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E.,
 Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 2235534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source
 1..706
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Layer and broiler"
 /db_xref="taxon:9031"
 /clone="CHSRT239b14"
 /sex="Male and female"
 /tissue_type="Chondrocytes isolated from growth plate cartilage"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSROBNI10"
 /notes="Vector: Bluescript II KS(+), Site 1: EcoRI, Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the Bluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 189 a 140 c 179 g 198 t
 ORIGIN

Query Match 18.6%; Score 166.8; DB 13; Length 706;
 Best Local Similarity 65.5%; Pred. No. 4e-37;
 Matches 262; Conservative 0; Mismatches 132; Indels 6; Gaps 1;

```

QY 7 GTGAACCAAGCCACCAAGAGCGTCGTGGGGCCCTCATCCCTCTGGCGAAAGTGTG 66
DB 27 GCGCTCAATAGGCAACATCGAAGAGCGGTGTGTCTGTCTCCCAATGCCAAGACATT 86
QY 67 TTGAAGCAGTGTAGAGATGTGACCTCTGCTTCTACAGAAACCAAGTGAATCTATCTG 126
DB 87 CTCACGCAATGTAAAGATGTGAGCTCTCTTCAGTACATGACGGCAAGCTTGAAGGCC 146
QY 127 TTTAATGCGCAAAAGAAAGAAAGTGTTCATCTTCACTTACCCGGTGTCTTCGTAAT 186
DB 147 TTCAAAGGCAACAAAGAGAAAGTGTATCTCACCCCAATACAGATGAATCTTCGTGTC- 205
QY 187 TCACACTTAGTCAATGATACCCCATGCTTCTTTATGATGCGGTTTGACCTGATAGTGAC 246
DB 206 -----AAAGGCAAGATCTATGCTGTCTTTATATATGCGTTTATTTGGTAAAGGG 260
QY 247 TGCACCATTGAAACCAATTTTGGCCCACTAATTAAGAAACCAATTCAGGCAAGCT 306
DB 261 TGCTTATTGAGAGAGCTCTTTCTCTGCTAATTAATCAAGAAAGATTCAGGCTGAG 320
QY 307 CCAGGTGTGTGCTGGGAGAGCAAGCTGTTTAAATGATCTTCAGAAAGAGAGGTGCC 366
DB 321 GCAGAGGTGTGCTGGGAGAGGAGAAAGCTTTAACTGACTTTCAACAGCGGAGAGGCC 380
QY 367 ATCGAATTGCCCACTGATGTATAAAGCTGCTGTGCTG 406

```

DB 381 ATCGAATTGGACAGACTGATGTCAAGCTGCTCTGATG 420

Search completed: December 16, 2003, 17:31:39
 Job time: 1850.09 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 15:10:30 / Search time 3348.76 Seconds

(without alignments)
3823.721 Million cell updates/sec

Title: US-09-864-291-5

Perfect score: 1709

Sequence: 1 MAVVQSHTEBRGALIPSGR.....ASHRSMTAQOETSLPTSSS 313

Scoring table:

BLOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPRO.spool/p/US09864291/runat.15122003.160858.23305/app.query.fasta.1.846
-DB=Genemb1 -OPMT=fastcap -SUFFIX=xge -MINMATCH=0.1 -LOOPCTL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.ccd1 -LIST=45
-LOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09864291.OCEN.1.1.3854.0runat.15122003.160858.23305 -NCPV=6 -ICPV=3
-NO_MMAP -LARGEOUBRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Genemb1:*
1: gb Da:*
2: gb Hcg:*
3: gb In:*
4: gb Om:*
5: gb Ov:*
6: gb Pat:*
7: gb Ph:*
8: gb Pl:*
9: gb Pr:*
10: gb Ro:*
11: gb Sts:*
12: gb Sy:*
13: gb Un:*
14: gb Vt:*
15: em Ba:*
16: em Fun:*
17: em Hum:*
18: em In:*
19: em Mu:*
20: em Om:*
21: em Or:*
22: em Ov:*
23: em Pat:*
24: em Ph:*
25: em Pl:*
26: em Ro:*
27: em Sts:*
28: em Un:*

29: em Vi:*
30: em Htg_hum:*
31: em Htg_liv:*
32: em Htg_other:*
33: em Htg_mus:*
34: em Htg_pln:*
35: em Htg_rod:*
36: em Htg_mam:*
37: em Htg_vrt:*
38: em By:*
39: em Htgo_hum:*
40: em Htgo_mus:*
41: em Htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1709	100.0	1413	6 AX359658	AX359658 Sequence
2	1036	60.6	2267	9 BC022546	BC022546 Homo sapi
3	1025	60.0	2266	9 BC022549	BC022549 Homo sapi
4	769	45.0	1001	6 AX359665	AX359665 Sequence
5	503	29.4	179222	2 BX296515	BX296515 Sus scrofa
6	503	29.4	222469	2 BX470149	BX470149 Danio rerio
7	475	27.8	1691	9 HSU79458	HSU79458 Human WM do
8	475	27.8	1820	9 BC007452	BC007452 Homo sapi
9	475	27.8	1835	9 AK057881	AK057881 Homo sapi
10	475	27.8	1912	9 BC010616	BC010616 Homo sapi
11	460	26.9	793	10 AF499026	AF499026 Rattus no
12	459.5	26.9	1885	6 AR220846	AR220846 Sequence
13	456	26.7	1802	10 MMT40826	MMT40826 Mus muscu
14	445	26.4	220895	9 HS250D10	HS250D10
15	445	26.0	2170	9 AB072784	AB072784 Macaca fa
16	424	24.8	222469	2 BX470149	BX470149 Danio rer
17	408.5	23.9	12964	2 BX296515	BX296515 Sus scrofa
18	408	23.9	1747	3 AY060828	AY060828 Drosophila
19	405	23.7	1205	5 BC053232	BC053232 Danio rer
20	392.5	23.0	168425	2 AC113593	AC113593 Mus muscu
21	392.5	23.0	224086	10 AC104325	AC104325 Mus muscu
22	377	22.1	253149	2 AC107527	AC107527 Rattus no
23	377	22.1	270171	2 AC132969	AC132969 Rattus no
24	332.5	19.5	110000	2 AC118467_0	AC118467 Mus muscu
25	321	18.8	454	6 BD027759	BD027759 Sequence
26	310	18.1	42845	2 AC017247	AC017247 Drosophila
27	310	18.1	175679	3 AC010025	AC010025 Drosophila
28	310	18.1	17590	2 AC006563	AC006563 Drosophila
29	310	18.1	297381	3 AB003540	AB003540 Drosophila
30	284.5	16.6	170781	2 AC092212	AC092212 Trypanoso
31	279	16.3	163542	2 AP001190	AP001190 Homo sapi
32	279	16.3	184371	9 AC104961	AC104961 Homo sapi
33	258	15.1	39782	2 AC138635	AC138635 Homo sapi
34	258	15.1	40476	2 AC140538	AC140538 Homo sapi
35	256.5	15.0	33899	3 CEY63D3A	CEY63D3A Caenorhab
36	256	15.0	148851	9 HS155D22	HS155D22 Human DNA s
37	254.5	14.9	195859	14 AF281817	AF281817 Thupaia he
38	249	14.6	159064	9 AC138512	AC138512 Homo sapi
39	249	14.6	174921	2 AC012266	AC012266 Homo sapi
40	245.5	14.4	130742	2 AC138601	AC138601 Mus muscu
41	245.5	14.4	187391	2 AC116738	AC116738 Mus muscu
42	244.5	14.3	148055	2 AC128178	AC128178 Rattus no
43	244.5	14.3	275392	2 AC095297	AC095297 Rattus no
44	243	14.2	72300	2 AC116403	AC116403 Mus muscu
45	243	14.2	167067	10 AC124489	AC124489 Mus muscu

RESULT 1

ALIGNMENTS

AX359658
 LOCUS AX359658 1413 bp DNA linear PAT 13-FEB-2002
 DEFINITION Sequence 4 from Patent WO0190185.
 ACCSSION AX359658
 VERSION AX359658.1 GI:18675409
 KEYWORDS
 ORGANISM
 SOURCE
 unclassified
 unclassified
 unclassified
 unclassified

REFERENCE
 1
 Oko, R. and Sutovsky, P.
 P132 sperm protein, sperm c-yes, oocyte cytoplasmic c-yes, and uses thereof
 Patent: WO 0190185-A 4 29-NOV-2001;
 QUEEN'S UNIVERSITY AT KINGSTON (CA) ; OREGON HEALTH SCIENCES UNIVERSITY (US)

FEATURES
 source
 1. 1413
 /organism="unclassified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 /note="mammalia"
 36. 977
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD23336.1"
 /db_xref="gi:18675410"
 /translation="MAVNSHTSRRLGLIPSGRSLVKQCEVDVLCFLQKPVBSYLFN
 CTXKGTPLTSTRYRYVPTVSHLVNDPMSFPMPPRLMSDCTTQGIIPAPNYIKGTIDAF
 PGGMBGQAVKLSFRKGLALEQMLKASALARKGIPLSAVTWPDTSGLYITTP
 GAACSSGTPCPAYIVYGPDPGYTVQPSYGTPEPQAGQKGYGAPPMGYPAGP
 VGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGV
 NEALPPAYEASAGNTAASHTASRMTAQCTSLPTSSS"

CDS
 36. 977
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD23336.1"
 /db_xref="gi:18675410"
 /translation="MAVNSHTSRRLGLIPSGRSLVKQCEVDVLCFLQKPVBSYLFN
 CTXKGTPLTSTRYRYVPTVSHLVNDPMSFPMPPRLMSDCTTQGIIPAPNYIKGTIDAF
 PGGMBGQAVKLSFRKGLALEQMLKASALARKGIPLSAVTWPDTSGLYITTP
 GAACSSGTPCPAYIVYGPDPGYTVQPSYGTPEPQAGQKGYGAPPMGYPAGP
 VGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGVGVPPGV
 NEALPPAYEASAGNTAASHTASRMTAQCTSLPTSSS"

BASE COUNT 377 a 363 c 369 g 304 t

ORIGIN

Alignment Scores:
 Pred. No.: 2e-79 Length: 1413
 Score: 1709.00 Matches: 313
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0

US-09-864-291-5 (1-313) x AX359658 (1-1413)

Cy 1 MetAlaValAlaInGlnSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
 Db 36 ATGGAGTGAACCGAGAGCCACCGAGAGCCGTCGTGGGGCCCTCATCCCTCTGGCGAA 95
 Cy 21 SerValIleuLysGlnCysGluAspValAlaAspLeuGlnLysProValGlnSer 40
 Db 96 AGTGTCTGAAGCAGTGAAGATGTGACCTCTGCTTCTACAGAAACCGATGGATCC 155
 Cy 41 TyrLeuPheAsnGlyThrIleuLysGlyThrIleuPheLeuSerTyrArgValAlaPhe 60
 Db 156 TATCTCTTAATGGCAAAAGAAAGAAAGCTTTTCTTCATCATACCGGGTGGTCTTC 215
 Cy 61 ValThrSerHisLeuValAlaAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db 216 GTGACTTCACACTTAAGTCAATGACCCCAATGCTTCTTTAATGATGCCCTTGGCCATAG 275
 Cy 81 SerAspCysThrIleGluGlnProIlePheAlaProAsnTyrIleuLysGlyThrIleGln 100
 Db 276 AGTGACTGACACTTAAGTCAATGACCCCAATGCTTCTTTAATGATGCCCTTGGCCATAG 335
 Cy 101 AlaAlaProGlyGlyGlyTyrGluGlnAlaAlaPheLysLeuSerPheArgIleGly 120
 Db 336 GCAAGCTCCAGGTGGTGGTGGGAGAGCAAGCTGTTTAAATATCTTCAGAGAAAGCA 395
 Cy 121 GlyAlaIleGluPheAlaGlnLeuMetValIleAlaAlaSerAlaAlaAlaArgGlyIle 140
 Db 396 GTGTCCATCGAATTTGCCCAACTGATGTGTAAGCTGCTCTGCTGCGCAGAGAAATTT 455

Cy 141 ProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrValPro 160
 Db 456 CCACTTGGAAGTGAATTAATCTGCTTGCACATTCAGAGCACTGATCAATTAATCTGCCA 515
 Cy 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGly 180
 Db 516 GGGGCTGACAGTGTCTCTCCACAGACCTTGTCCAGCATATCAATGATGATCTATGGA 575
 Cy 181 ProProProProGlyTyrThrValGlnProGlyGlyTyrGlyLysThrProProGlyTyr 200
 Db 576 CCCCACACACAGATATATACATGTCACAGGGGATATATGAACTCCACAGAAAGATAT 635
 Cy 201 GlyAlaGlnProGlyGlyTyrGlyAlaProPheMetGlyTyrGlyAlaProProValGly 220
 Db 636 GAGGCCCAACAGGGGGATATGAGAGCCCACTATGGAATATGAGAGCCCGGCTGTGGA 695
 Cy 221 TyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProProGly 240
 Db 696 TATGAGTCCACCTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACTGGG 755
 Cy 241 GlyTyrGlyAlaProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProPro 260
 Db 756 GATATGAGAGCCCACTGGGGATATGAGTCCCACTGGGGATATGAGTCCCACTGGG 815
 Cy 261 GlyGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProAlaGlyAlaGlnValAlaLeu 280
 Db 816 GGGGATATGAGAGCCCACTGGAGATATGAGAGCCCACTGGAGATATGAGAGCCCTA 875
 Cy 281 ProProAlaTyrGlyAlaProSerAlaGlyAsnThrAlaAlaSerHisAspSerMetThr 300
 Db 876 CCCCCTGCATATGAGTCCATCTGCTGGAATAACAGCTCCCTCAAGATCATATGACA 935
 Cy 301 AlaGlnGlnGlnThrSerLeuProThrThrSerSerSer 313
 Db 936 GCTCAGAGAGAGACTCTCTCCCACTCACTCATCTTCT 974

RESULT 2
 BC022546 2267 bp mRNA linear PRI 04-FEB-2002
 LOCUS
 DEFINITION Homo sapiens, similar to RIKEN cDNA 4930521I23 gene, clone
 WGC:26816 IMAGE:4811804, mRNA, complete cds.
 ACCESSION BC022546
 VERSION BC022546.1 GI:18490706
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2267)
 Struhsberg, R.
 Submitted (01-FEB-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: gcgabs-remail.nih.gov
 Tissue Procurement: Mikhail Palkovite, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Tohyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LANT)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome
 Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-sngc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
 R. M.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LANT at: <http://image.lnl.gov>
 Series: IRAX Plate: 32 Row: J Column: 20

CDS

```

/class type="Brain, hippocampus"
/clone_id="NH_MGC_95"
/lab host="DH10B"
/notes="Vector: pBluescript"
32..634
/codon_start=1
/product="similar to RIKEN CDNA 4930521123 gene"
/protein_id="AAH22549.1"
/db_xref="GI:18490711"
/translation="MAVNSSTHNRRLIPNGSLIKSPNVELSPRSGSNVS
GRKGTLPITSRYIPITSGISIDPHLSFPMPLMTLTLTBOVPANPTKGTQAA
PYGMEGQATKLIVRNGALIFPAQILVKAASAAAGPPLTLNDWFSNGIYITGK
GICALHRLVLLSMEPHLDMEPHLPTEPHLDMEBPNP"

```

BASE COUNT 709 a 477 c 468 g 612 t
ORIGIN

Alignment Scores:

Pred. No.: 3,13e-44 Length: 2266
Score: 1025.00 Matches: 200
Percent Similarity: 73.10% Conservative: 31
Best Local Similarity: 63.29% Mismatches: 71
Query Match: 59.98% Indels: 15
DB: Gaps: 4

US-09-864-291-5 (1-313) x BC022549 (1-2266)

```

Oy 1 MetAlaValaInGlnSerHisThrGlnSerArgArgIValaLeuIleProSerGlyGlu 20
Db 32 ATGGCGGTGATTCAGAGCCACACCGAAGACCGCCGAGCCCTCATTCCTTAACGGTGA 91
Oy 21 SerValleuLysGlnCysGluAspValaAspLeuCySpheLeuGlnLysProValaGlnSer 40
Db 92 AGTCTCTTGAAGCGGCTCCGAAATGTGAGGCTCTCTCCACAGCATTCAGAAAGCTCA 151
Oy 41 TyrLeuPheAsnGlyThrLysLysGlyThrLeuPheLeuThrSerTyrArgValaPhe 60
Db 152 AATGCTTAGGGTGAAGAAAGACAGAAATGTTCTTCATCTCAACCGGATGATTTTC 211
Oy 61 ValThrSerHisLeuValaAsnAspProMetLeuSerPheMetProPheGlyLeuMet 80
Db 212 ATTAATCTCAATGCTCATCATGATTCATGTTGTTCTTTATGATGCCATTTGATCTGATG 271
Oy 81 SerAspCyThrIleGluGlnProIlePheAlaProAsnTyrIleLysGlyThrIleGln 100
Db 272 AGCAACCTCATCTTGAACACACAGATATTGCTGAACTTCATTAAAGGAATATTGAG 331
Oy 101 AlaAlaProGlyGlyGlyTTPGluGlyGlnAlaValPheLysLeuSerPheArgLysGly 120
Db 332 GCAGCTCATATGCTGCTGGAGAGACAACTTAAATTAGTCTTCAGAAATGGA 391
Oy 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaSerAlaAlaArgLysIle 140
Db 392 GGTGCTATTGAATTTGCGCATGTATGTGAAGAGCTGCTGCTGCTGCGCCGAGAAATT 451
Oy 141 ProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrValPro 160
Db 452 CCACTTAGAACCTTAATGACGCTTCAAGCTCTTAAGGAATTAATGATTAATGACG-AAA 510
Oy 161 GlyAlaAlaValCysSerSerGlnThrProCySpProAlaTyrProIleValIleTyrGly 180
Db 511 GGG---AATATGTGACCTCCACAGATGCTCTGT-----TCAGTAATGTCTATGGA 558
Oy 181 ProProProProGlyTyrThrValGlnProGlyGlyLysTyrGlyThrProProGluGlyTyr 200
Db 559 GCCCACTGAGAGA-----TATGAGCCCACTCCGAGATTC 597
Oy 201 GlyAlaGlnProGlyGlyTyrGlyValaProPheMetGlyTyrGlyValaProProValGly 220
Db 598 GGAAGCCCACTGAGAGATATGAGAGCCCAACCGATGAGAAATGAAAGCCCGCTGTGGA 657
Oy 221 TyrGlyValaProProGlyGlyTyrGlyValaProProGlyGlyTyrGlyValaProProGly 240
Db 658 TACAGAGCTCACTGTGTGATATGAGAGCCCACTCTTGATATGAGAGCCCACTGCA 717

```

```

Oy 241 GlyTyrGlyValaProProGlyGlyTyrGlyValaProProGlyGlyTyrGlyValaProPro 260
Db 718 GATATATGAGAGCCCACTTATGATATGAGAGCCCACTTATGATATGAGAGCCCACTT 777
Oy 261 GlyGlyTyrGlyValaProProAlaGlyTyrGlyValaProProAlaGlyValaGlnLeu 280
Db 778 CTGGATATGAGAGCCCACTTCTGATATGAGAGCCCACTTCTGAGAAATGAAAGCCCG 837
Oy 281 ProProAlaTyrGluAlaProSerAlaGlyAsnThrAlaAlaSerHisArgSerMetThr 300
Db 838 CTGGAGGATATCAGAGCTTCACTGCTGATATCAGAGAGCCCTCAGAAATCTTACAGCA 897
Oy 301 AlaGln-----GlnGlnThrSerLeuProThrThrSerSerSer 313
Db 898 GCCAGGCTCTGTAAGAGAGGCTTCTTCTCCCTGCTGCTCTTCT 945

```

RESULT 4

AX359665

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 1.8e-31 Length: 1001
Score: 769.00 Matches: 148
Percent Similarity: 68.60% Conservative: 18
Best Local Similarity: 61.16% Mismatches: 62
Query Match: 45.00% Indels: 14
DB: Gaps: 4

US-09-864-291-5 (1-313) x AX359665 (1-1001)

```

Oy 75 MetProPheGlyLeuMetSerAspCyThrIleGluGlnProIlePheAlaProAsnTyr 94
Db 1 ATGCCATTATATCTGATGACGAACTCATCTTAAACACACAGATTTTTCGCAAACTTC 60
Oy 95 IleLysGlyThrIleGlnAlaAlaProGlyGlyGlyTTPGluGlyGlnAlaValPheLys 114
Db 61 ATTAAGGGAATTAATTCAGAGGCTCCATATGTGTGCTGGAGAGGACAAAGCTACTTTAAA 120
Oy 115 LeuSerPheArgLysGlyGlyValaIleGluPheAlaGlnLeuMetValLysAlaSer 134

```

Db 121 TTAGTCTTCAGAAATGAGATGCAATGAAATTTGCCAGTTGATGTAAGCAAGCTCTT 180
 QY 135 AAlaAlaAarglyIleProLeuGlySerValAsnTyTrpPheAspThrSerGlyLeu 154
 Db 181 GCTGTGGCCGAGAGATTTCCACTTGAACCTTAAATGACTGTTCAGCTCTATGAGAAAT 240
 QY 155 TTTleIleTherValProGlyAlaAlaValCysSerSerGlnThrProCysProAlaTy 174
 Db 241 TATGTAAATTCCTGGGAGAGG---AATATGTCACCTCCACAGATGCTTGT----- 288
 QY 175 ProIleValIleTyrglyProProProGlyTyTrpThraValGlnProGlyGlyTy 194
 Db 289 TCAGTATGTCTATGAGGAGGCCCACTGACAGA-----TATGGA 327
 QY 195 ThrProProGlyIleTyrglyAlaGlnProGlyTyTrpThraValAlaProProMetGlyTy 214
 Db 328 GCCCACCCTCCGAAATAGGAGGCCCACTGACAGAAATAGAGCCCAACCCGTAGAGAAAT 387
 QY 215 GYAlaAProProValGlyTyrglyValAlaProProGlyTyTrpThraValAlaProProGlyTy 234
 Db 388 GAAGGCCCGCTGTGGATACAGAGCTCACCTGTGGAATATGAGGCCCACTCTTGGGA 447
 QY 235 TyrglyValProProGlyGlyTyrglyAlaProProGlyTyTrpThraValAlaProProGlyTy 254
 Db 448 TAGGAGGCCCACTGACAGATATGAGGCCCACTCTAGAAATATGAGGCCCACTCTT 507
 QY 255 GYTyrglyValAlaProProGlyTyTrpThraValAlaProProAlaGlyTyTrpThraValAlaProPro 274
 Db 508 GATATAGAACCCCACTCTCGATATGAGGCCCACTCTCGAATATGAGGCCCACT 567
 QY 275 AlaGlyAlaGlnAlaLeuProProAlaTyrglyAlaAlaProSerAlaGlyAlaAla 294
 Db 568 GCAGAAATATAGGCCCGCTGCGGATACAGAGCTCACCTGTGATATGAGGCCCAAG 627
 QY 295 SerHisArgSerMetThrAlaGln-----GlnGlnThrSerLeuProThrSer 311
 Db 628 CTCGAGAAATCTACAGAGGCCCAAGCTCTGAAAGAGAGCTCTCTCCCTGCTGCC 687
 QY 312 SerSer 313
 Db 688 TCTTCT 693
 RESULT 5
 LOCUS BX296515 179222 bp DNA linear HTG 26-MAY-2003
 DEFINITION Sus scrofa clone PigR-121D21. *** SEQUENCING IN PROGRESS ***, 6
 unsorted pieces.
 ACCESSION BX296515
 VERSION BX296515.6 GI:31076160
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 1 (bases 1 to 179222)
 Tracey, A.
 Direct Submission
 Submitted (25-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 On May 26, 2003 this sequence version replaced gi:31043704.
 COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BT121D21
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 177502 bases at least Q40
 Consensus quality: 177971 bases at least Q30

Consensus quality: 178283 bases at least Q20
 Insert size: 178722; sum-of-contigs
 Insert size: 167315; 12.1% error; agarose-fp
 Quality coverage: 7.82x in Q20 bases; sum-of-contigs Quality
 coverage: 8.59x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 5850: contig of 5850 bp in length
 * 5851 gap of 100 bp
 * 5951 74264: contig of 68314 bp in length
 * 74265 74364: gap of 100 bp
 * 74365 84656: contig of 10292 bp in length
 * 84657 84756: gap of 100 bp
 * 84757 108493: contig of 23737 bp in length
 * 108494 108593: gap of 100 bp
 * 108594 114454: contig of 5861 bp in length
 * 114455 114554: gap of 100 bp
 * 114555 179222: contig of 64668 bp in length.
 Location/Qualifiers
 1..179222
 /organism="Sus scrofa"
 /mol_type="genomic DNA"
 /db_xref="taxon:9823"
 /clone="PigR-121D21"
 /clone_1lb="PigR"
 1..5850
 /note="assembly_fragment:02838
 fragment_chain:1"
 5951..74264
 /note="assembly_fragment:01479
 fragment_chain:1"
 74265..84656
 /note="assembly_fragment:01113
 fragment_chain:1"
 84757..108493
 /note="assembly_fragment:00846
 fragment_chain:1"
 108594..114454
 /note="assembly_fragment:01058
 fragment_chain:2"
 114555..179222
 /note="assembly_fragment:02416
 fragment_chain:2"
 BASE COUNT 44892 a 44880 c 44117 g 44831 t 502 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,05e-15 Length: 179222
 Score: 503.00 Matches: 101
 Percent Similarity: 58.56% Conservative: 5
 Best Local Similarity: 55.80% Mismatches: 33
 Query Match: 29,434 Indels: 42
 DB: 2 Gaps: 4
 US-09-864-291-5 (1-313) x BX296515 (1-179222)
 QY 175 ProIleValIleTyrglyProProProGlyTyTrpThraValGlnProGlyGlyTy 194
 Db 55265 CCAAGTGTGTCTATGAGACCCCTACAGATATGAGGCCCAACCAAGATATTA 55324
 QY 195 ThrProProGlyGlyTyrglyAlaGlnProGlyTyTrpThraValAlaProProMetGlyTy 214
 Db 55325 GCCCACCCTCCAGATATGAGGCCCACTCTCGAATATGAGGCCCACTCTCAGAGATAC 55384
 QY 215 GYAlaAProProValGlyTyrglyValAlaProProGlyTyTrpThraValAlaProProMetGlyTy 229

Db 55385 GAGAGCCCACTCCAGATACGAGAGCCCACTCCAGATACGAGAGCCCACTCCAGAGAG 55444
 QY 230 -----ValProGlyGlyTyrGlyValPro----- 238
 Db 55445 CTGAGAGCCCACTCCAGATACGAGAGCCCACTCCAGATACGAGAGCCCACTCCAGAG 55504
 QY 239 -----ProGlyGlyTyrGly-----AlaProPro 246
 Db 55505 AGATATGACGCGCTCCAGATACGAGAGCCCACTCCAGATACGAGAGCCCACTCCAGAG 55564
 QY 247 GlyGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValPro 266
 Db 55565 GCAGATATGAGAGTCCCTCCAGATACGAGAGCCCACTCCAGATACGAGAGCCCACTCCAGAG 55624
 QY 267 ProAlaGlyTyrGlyAlaProPro----- 274
 Db 55625 CCTCCAGATATGAGAGTCCCTCCAGATACGAGAGCCCACTCCAGATACGAGAGCCCACTCCAGAG 55684
 QY 275 -----AlaGlyAangluAlaLeuProProAlaTyrGlyAlaProSerAlaGlyAanThr 292
 Db 55685 ACAACAGCTGAAATGAGAGCCCACTCCAGATACGAGAGCCCACTCCAGATACGAGAGAG 55744
 QY 293 AlAlaSerAlaGlySerMetThrAlaGlnGlnGlnThrSerLeuProThrThrSerSer 312
 Db 55745 GCTGCTCTCCCAATCTGTGAGAGCCCACTCCAGATACGAGAGCCCACTCCAGATACGAGAG 55804
 QY 313 Ser 313
 Db 55805 TCT 55807
 RESULT 6
 Bx470149 222469 bp DNA linear HTG 07-MAY-2003
 LOCUS Bx470149
 DEFINITION Dario rerio clone DKRY-242K7, *** SEQUENCING IN PROGRESS ***, 50
 unoriented pieces.
 ACCESSION Bx470149
 VERSION Bx470149.3 GI:30424228
 KEYWORDS HTG; HTGS. PHASE1.
 SOURCE Dario rerio (zebrafish)
 ORGANISM Dario rerio
 Burkholderia, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
 Actinopterygii, Neopterygii, Teleostei, Osteichthyes,
 Cypriniformes, Cyprinidae, Danio.
 1 (bases 1 to 222469)
 Burton, J.
 Direct Submission
 Submitted (06-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On May 7, 2003 this sequence version replaced gi:30387077.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zfish-help@sanger.ac.uk
 ----- Project Information
 Center project name: zK242K7
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 197958 bases at least Q40
 Consensus quality: 206715 bases at least Q30
 Consensus quality: 212351 bases at least Q20
 Insert size: 217569; sum-of-contigs
 Insert size: 165837; 6.2% error; agarose-fp
 Quality coverage: 2.48x in Q20 bases; sum-of-contigs Quality
 coverage: 3.95x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 50 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 6656: contig of 6656 bp in length
 * 6657: gap of 100 bp
 * 6757: contig of 3499 bp in length
 * 10256: gap of 100 bp
 * 10356: contig of 2866 bp in length
 * 10356: gap of 100 bp
 * 13324: contig of 6612 bp in length
 * 13324: gap of 100 bp
 * 19935: gap of 2562 bp in length
 * 20036: gap of 100 bp
 * 22598: contig of 100 bp
 * 22598: gap of 100 bp
 * 22698: contig of 5872 bp in length
 * 22698: gap of 100 bp
 * 28570: contig of 9296 bp in length
 * 28570: gap of 100 bp
 * 37966: gap of 100 bp
 * 38066: contig of 3075 bp in length
 * 41141: gap of 100 bp
 * 41241: contig of 3866 bp in length
 * 45109: gap of 100 bp
 * 45209: contig of 2522 bp in length
 * 47331: gap of 100 bp
 * 47831: contig of 4054 bp in length
 * 51885: gap of 100 bp
 * 51885: contig of 3873 bp in length
 * 51885: gap of 100 bp
 * 55957: gap of 100 bp
 * 55958: contig of 4541 bp in length
 * 60499: gap of 100 bp
 * 60599: contig of 3043 bp in length
 * 63641: gap of 100 bp
 * 63741: contig of 8933 bp in length
 * 63742: gap of 100 bp
 * 72675: contig of 2828 bp in length
 * 72775: gap of 100 bp
 * 75603: contig of 2360 bp in length
 * 75703: gap of 100 bp
 * 78063: contig of 100 bp
 * 78163: gap of 100 bp
 * 84486: contig of 6323 bp in length
 * 84486: gap of 100 bp
 * 84586: contig of 2966 bp in length
 * 87554: gap of 100 bp
 * 87654: contig of 4834 bp in length
 * 92488: gap of 100 bp
 * 92588: contig of 3517 bp in length
 * 96105: gap of 100 bp
 * 96205: contig of 5275 bp in length
 * 101480: gap of 100 bp
 * 101580: contig of 3400 bp in length
 * 101580: gap of 100 bp
 * 105080: contig of 2924 bp in length
 * 108004: gap of 100 bp
 * 108104: contig of 5046 bp in length
 * 113152: gap of 100 bp
 * 113152: contig of 3334 bp in length
 * 113252: gap of 100 bp
 * 116586: contig of 2661 bp in length
 * 116586: gap of 100 bp
 * 119347: contig of 3679 bp in length
 * 119447: gap of 100 bp
 * 123126: contig of 8041 bp in length
 * 123226: gap of 100 bp
 * 131267: contig of 4827 bp in length
 * 131267: gap of 100 bp
 * 136194: contig of 100 bp
 * 136294: gap of 100 bp
 * 136294: contig of 3189 bp in length
 * 139483: gap of 100 bp
 * 139483: contig of 2293 bp in length
 * 141875: gap of 100 bp
 * 141875: contig of 2147 bp in length
 * 141976: gap of 100 bp
 * 144123: contig of 2746 bp in length
 * 144223: gap of 100 bp
 * 146968: contig of 100 bp
 * 147068: gap of 100 bp
 * 149399: contig of 2331 bp in length
 * 149400: gap of 100 bp

```

* 149500 153963: contig of 446 bp in length
* 153964 154063: gap of 100 bp
* 154064 156579: contig of 2516 bp in length
* 156580 156679: gap of 100 bp
* 156680 158928: contig of 2249 bp in length
* 158929 159028: gap of 100 bp
* 159029 164542: contig of 5514 bp in length
* 164543 172054: contig of 7412 bp in length
* 172055 172154: gap of 100 bp
* 172155 175210: contig of 3056 bp in length
* 175211 175310: gap of 100 bp
* 175311 179379: contig of 4069 bp in length
* 179380 179479: gap of 100 bp
* 179480 182698: contig of 3219 bp in length
* 182699 182798: gap of 100 bp
* 182799 186779: contig of 3981 bp in length
* 186780 186879: gap of 100 bp
* 186880 195626: contig of 8747 bp in length
* 195627 195726: gap of 100 bp
* 195727 199396: contig of 3670 bp in length
* 199397 199496: gap of 100 bp
* 199497 206412: contig of 6916 bp in length
* 206413 206512: gap of 100 bp
* 206513 209619: contig of 3107 bp in length
* 209620 209719: gap of 100 bp
* 209720 219083: contig of 9364 bp in length
* 219084 224659: contig of 3286 bp in length.
* 219184 Location/Qualifiers
1. 222469

```

```

FEATURES
Source
1. 222469

```

```

/misc_feature
/note="assembly_fragment:01359
fragment_chain:1"
6757..10255
/note="assembly_fragment:01385
fragment_chain:1"
10356..13223
/note="assembly_fragment:00592
fragment_chain:1"
13324..19935
/note="assembly_fragment:00237
fragment_chain:1"
20036..32597
/note="assembly_fragment:00949
fragment_chain:1"
22698..28569
/note="assembly_fragment:00615
fragment_chain:1"
28670..37965
/note="assembly_fragment:01306
fragment_chain:2"
38066..41140
/note="assembly_fragment:00487
fragment_chain:2"
41241..45108
/note="assembly_fragment:00232
fragment_chain:2"
45209..47730
/note="assembly_fragment:00935
fragment_chain:2"
47831..51884
/note="assembly_fragment:01415
fragment_chain:2"
51985..55857
/note="assembly_fragment:00878
fragment_chain:3"
55958..60498

```

```

/misc_feature
/note="assembly_fragment:01208
fragment_chain:3"
60599..63641
/note="assembly_fragment:01266
fragment_chain:3"
63742..72674
/note="assembly_fragment:01329
fragment_chain:3"
72775..75602
/note="assembly_fragment:01183
fragment_chain:4"
75703..78062
/note="assembly_fragment:01179

```

Alignment Scores:

Pred. No.:	1 296-15	Length:	222469
Score:	503.00	Matches:	101
Percent Similarity:	58.56%	Conservative:	5
Best Local Similarity:	55.80%	Mismatches:	33
Query Match:	29.43%	Indels:	42
DB:	2	Gaps:	4

US-09-864-291-5 (1-313) x BK470149 (1-222469)

```

QY 175 ProIleValIleTyrGlyProProProGlyTyrThrValIGlnProGlyIuTyrGly 194
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 196921 CCAAGTGTGTCTATGAGACCCCTACCAAGATATGAGGCCACACAGATATATAGA 196980

QY 195 ThrProProGlyGlyTyrGlyValaGlnProGlyGlyTyrGlyValaProProMetGlyTyr 214
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 196981 GCCCCACCTCCAGATATACGAGACCCCACTCCAGATATAGAGGCCCACTCCAGATATAC 197040

QY 215 GAlaAProProValGlyTyrGlyValaProProGlyGlyTyrGlyTyrGlyTyrGlyTyrGly 229
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 197041 GAGGCCCACTCCAGATATACGAGACCCCACTCCAGATATAGAGGCCCACTCCAGATATAC 197100

QY 230 -----ValProProGlyGlyTyrGlyValaPro----- 238
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 197101 CTCGAAAGCCCACTCCAGATATAGAGGCCCACTCCAGATATAGAGGCCCACTCCAGATATAC 197160

QY 239 -----ProGlyGlyTyrGlyTyrGlyTyrGlyTyrGlyTyrGlyTyrGlyTyrGlyTyrGly 246
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 197161 AGATATAGAGGCCCTCCAGATATAGAGGCCCTCCAGATATAGAGGCCCTCCAGATATAGAGGCC 197220

QY 247 GAlaGlyTyrGlyValaProProGlyGlyTyrGlyValaProProGlyGlyTyrGlyValaPro 266
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 197221 GCGAGATATAGAGTCCCATCTCGGAGATATAGAGGCCCACTCCAGATATATATATATATATAT 197280

QY 267 ProAlaGlyTyrGlyValaProPro----- 274
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 197281 CCTCCAGATATAGAGGCCCTCCAGATATAGAGGCCCTCCAGATATAGAGGCCCTCCAGATATAGAGGCC 197340

QY 275 -----AlaGlyValaGlnIuValaLeuProProAlaTyrGlyValaProSerAlaGlyAanthr 292
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 197341 ACNAAGCTGGAATATGAGCCCACTTACGATATAGAGGCCCACTTACGGAATATAGAGGCCCACTTACG 197400

QY 293 AlaAlaSerHisArgSerMetThrAlaGlnGlnGlnIuThrSerLeuProThrThrSerSer 312
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 197401 GCGGCTCTCCAGAAATCTGTGCAAGGCCAGGAGGCTTCTCTCCCTTACCTCATCT 197460

QY 313 Ser 313
|||
Db 197461 TCT 197463

```

```

RESULT 7
LOCUS HSU79458 1691 bp mRNA linear PRI 01-FEB-1999
DEFINITION Human WW domain binding protein-2 mRNA, complete cds.
ACCESSION U79458
VERSION U79458.1 GI:4205085
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```


Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 contact: amadan@systemsbio.org
 Anup Madan, Rachel Dickhoff, Jessica Fahy, Stephanie Ford, Julia Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: RML Plate: 10 Row: b Column: 24
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency OMF analysis.

FEATURES

source

Location/Qualifiers
 1..1820
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:4586 IMAGE:3140484"
 /tissue_type="Placenta, choriocarcinoma"
 /clone_1b="N1H MGC 21"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 27..812
 /codon_start=1
 /product="Similar to WW domain binding protein 2"
 /protein_id="AAH07452.1"
 /db_xref="GI:1393601"
 /translation="MALNNHSGGGVYVNTSTIMSDHYELTPNDKVPBAPK
 TKGTAVLTPRVITLTKSKDMOSPMFPTMKDCETKOPFGNATIKGTAKAG
 GMSGASTLTATAGALBPGRMLQVASQSRGSPSAGVSYMPGSAVYPPVA
 NGMYCPGPPPPPPPPPPPPPPMDAGMVGPPPPPPPPPPPPPPVSGPVPSTPA
 AAKAAEAASAYVNPNGPHNYMPTSGPPPPPPPPPPPPEDKTKQ"

BASE COUNT

373 a 603 c 451 g 393 t

ORIGIN

Alignment Scores:
 Pred. No.: 3,52e-16 Length: 1820
 Score: 475.00 Matches: 132
 Percent Similarity: 47.088 Conservative: 37
 Best Local Similarity: 36.778 Mismatches: 113
 Query Match: 27.798 Indels: 77
 DB: 9 Gaps: 16

US-09-864-291-5 (1-313) x BC007452 (1-1820)

QY 1 MetcAlaValaAmgIserrHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
 DB 27 ATGGCGCTCAACAGATCACTCGAGGCGC--GGCGAGTGTATCTCAATAACACCGAG 83
 QY 21 SerValIleuLysGlnCyseGluSerValaLeuCysePheLeuGlnLysProValaGluSer 40
 DB 84 AGCATCTTAATGCTCATGATCACTGGAACCTCACTTAATGATGACATGAAGAACTGCCA 143
 QY 41 TyrLeuPheAmgIlyThrLysLysGlyThrLeuPheLeuThrSerTyrArgValaPhe 60
 DB 144 GAAGCTTCAAGAGGACCAAGAAAGCACTGCTCACTTACCCCTTACCGGGTCACTTT 203
 QY 61 ValThrSerHisLeuValaAmgAPProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 204 CTGTCCAGAGGC-----AAGGATGCGACGAGCTCTTCATGATGCACTTTATCTCATG 257
 QY 81 SerAPCyseThrIleGlnIleProIlePheAlaProAmgTyrIleLysGlyThrIleGln 100
 DB 258 AAAGCTGTGACATCAAGAGCCCGATTGTGCAACTACATCAAGGAAACAGAGAG 317
 QY 101 AlaAlaProGlyGlyGlyTyrGlnGlnAlaValaPheLysLeuSerPheArgLysGly 120
 DB 318 GCGGAAGGCGGAGGAGTGGCTGGAGAGCTCTGCTTCTTCAACTTCACTTCAAGGCAAGG 377

QY 121 GtAlaAlaIleGluPheAlaGlnLeuMetValIlyAlaAlaSerAlaAlaArgGly---- 139
 DB 378 GCGGCACTTGAAGTGGACAGCGGATGCTCCAGAGGATCTCAAGCTCCAGAGTAA 437
 QY 140 IleProLeuGlySerValaAmgTyrTrpPheAPThrSerGlyLeuTyrIleThrVal 159
 DB 438 GTCCCAAGTGAAGCATATGCTTACATCAAGCCAGGGGGCTATATGTATATCCCCG 497
 QY 160 ProGlyAlaAlaValaCyseSerSerGlnThrProGly---ProAlaTyrProIleValIle 178
 DB 498 CCA-----GTGCGCAATGAAATGATACCTCTGCTCTGCTACCTCC----- 539
 QY 179 TyrGlyProProProProGlyTyrThrValaGlnProGlyGluTyrGlyThrProProGlu 198
 DB 540 TATCCACCGCCCGCCCACTGAATTC-----TATCCAGAACCCCGCAG 581
 QY 199 GtYTrGtAlaAlaGlnProGlyGlyTyrGlyAlaProPro-----MetGlyTyr 214
 DB 582 ATGAGCGGGGCGCAG-----GATAGGTGAGAGCCCAACAGCGCCCTACCTGGGCGC 635
 QY 215 GtAlaAProProVala---GtYTrGtAlaPro----- 224
 DB 636 ATGGAACCTCGGTCAAGCGCGCCGATGCTCCCTCACTCTGACCGGAAGCAAGGCG 695
 QY 225 -----Pro 225
 DB 636 GCAGAGACGCGCGCAGCGCTCATTAACAACCGAGCAATCTCAACAGCTCAATGCTCC 755
 QY 226 GtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTr 236
 DB 756 ACGAGCCAGCGCGCGCAGCTCCCTCACTCAACCGGAAGATGAAGAACCAAGTAAAGCC 815
 QY 237 -----ValProProGlyGly---TyrGtAlaProPro----- 246
 DB 816 CTCTGCTCTCTCTGCTCTCCACCTCACTCTCACTCACTCACTCACTCACTCACTCACT 875
 QY 247 GtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTr 262
 DB 876 CTGGGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 935
 QY 263 TyrGtAlaAProPro-AlaGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTrGtYTr 282
 DB 936 TACCAAGAACTGACATTTGGAGACTTGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 992
 QY 282 AlaTyrGtAlaAProSerAlaGtYTrAlaAlaSerHisArgSerMetThr 300
 DB 993 CAGCTTCCCATGCGCAGCCCGGAGCCCAAGTGTGCTGCGAGGTACCTCTCTCACC 1047

RESULT 9

AK057881 1835 bp mRNA linear PRI 31-OCT-2001
 LOCUS DEFINTION Homo sapiens CDNA PJ25152 file, clone CB07790, highly similar to Human WW domain binding protein-2 mRNA.

AK057881.1 GI:16553852
 ACCESSION o11go capping; file (full insert sequence).
 VERSION
 KEYWORDS
 NAMES
 SOURCE

ORGANISM

Homo sapiens (human)
 Homo sapiens
 BUKARYOTA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
 Fukuzumi,Y., Fujimori,Y., Komiyama,M., Suzuki,Y., Hata,H.,
 Nakagawa,K., Mizuno,S., Morinaga,M., Kawamura,M., Sugiyama,T.,
 Irie,R., Otsuki,T., Sato,H., Nishikawa,T., Sugiyama,A.,
 Kawakami,B., Nagai,K., Isegai,T. and Sugano,S.
 NEBO human cDNA sequencing project
 Unpublished

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

2 (bases 1 to 1835)
 Sugano,S. and Suzuki,Y.
 Direct Submission
 Submitted (24-OCT-2001) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure, Human Genome

COMMENT

Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
 (E-mail: cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
 Fax: 81-3-5449-5416)
 NRO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction and 5'-end one pass sequencing: Institute of Medical
 Science, University of Tokyo, Laboratory of Genome Structure, Human
 Genome Center; 3'-end one pass sequencing: RAB; clone selection for
 full insert sequencing: RAB and Helix Research Institute.

FEATURES

SOURCE

1..1835
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CBR07790"
 /tissue_type="brain"
 /clone_id="CBR"
 /note="Cloning vector: pMB18SFL3"
 BASE COUNT 385 a 604 c 455 g 391 t
 ORIGIN

Alignment Scores:

Score: 3.55e-16 Length: 1835
 Percent Similarity: 475.00 Matches: 132
 Best Local Similarity: 47.08% Conservative: 37
 Query Match: 36.77% Mismatches: 113
 Indels: 77
 Gaps: 16

US-09-864-291-5 (1-313) x AK057881 (1-1835)

1 MetAlaValangInSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
 29 ATGGCGCTCAAGAAATCACTCGAGGCG--GGCGAGGATGCTGCAATACACCGAG 85
 21 SerValleuYsgInCygluAspValAspleuCySpheLeuGlnYsProValGluSer 40
 86 AGCATCTATGTCCTATGATCATCGTGAACCTCATTCATGACATGACATGAGAACGTC 145
 41 TyrLeuPheAsnGlyThrValGlyGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
 146 GAACCTTCAGAGGACCAAGAAAGGCACTGTCTACCTTACCGGCTCACTTT 205
 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 206 CTGTCCAGGCG-----AAGATGCCATGACGCTCTTCATGATGCCATTTTATCTCATG 259
 81 SerAspCyThrIleGluGlnProIlePheAlaProAsnTyrIleYsgIleThrIleGln 100
 260 AAAGACTGTGATCAACACGACCGCGTATTTGGTCAACTCATCAACAGGAAACAGTGAG 319
 101 AlaAlaProGlyGlyGlyTyrGluGlnAlaValPheYsLeuSerPheArgGlyGly 120
 320 GCGGAGGCGGAGGCTGCGAGGCTCTCTCTTCATCAAGTTGACTTTACGCGCAGG 379
 121 GlyAlaIleGluPheAlaGlnLeuMetValYsAlaAlaSerAlaAlaArgGly--- 139
 380 GCGCGCATGATGCTCGACACGCGATGCTCCAGGTGCGATCTCAAGCTCCAGGGGTAA 439
 140 IleProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuYrIleIleThrVal 159
 440 GTCCCAATGAGGCTATGCGCTATCACTTACATGACGCGGCGCTATGCTATGCCCG 499
 160 ProGlyAlaAlaValCysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
 500 CCA-----GTCCCAATGAGATGTATCCCTGCGCTCTGCGCTACCC----- 541
 179 TyrGlyProProProGlyTyrThrValGlnProGlyGlyGlyTyrGlyThrProGlu 198
 542 TATCACCGCGCCCACTGAGTTC-----TATCACGAGACCCCGCATG 583
 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProPro-----MetGlyTyr 214

584 ATGACGCGGCGCATG-----GGATACGTGACAGCCCCACACCGCCTTACCTGGGCC 637
 215 GlyAlaProProVal---GlyTyrGlyValPro----- 224
 638 ATGAACTCTCGATCAAGCGCGCCGATGTCCTTCATCTCTGACCGGACCGAGCGAGCC 697
 225 -----Pro 225
 698 GCAGAGACGACCGCGCGCTTATCAACCCAGGCAATCTTCACAAAGTCTATATGCC 757
 226 GlyGlyTyrGlyValProProGlyGlyTyrGly----- 236
 758 ACAGACGACGCGCGCGCTTATCAACCCAGGCAATCTTCACAAAGTCTATATGCC 817
 237 -----ValProProGlyGly---TyrGlyAlaProPro----- 246
 818 CTCTGCT 877
 247 GlyGlyTyrGlyValProProGlyGlyTyrGlyAlaProProGly-----Gly 262
 878 CTGGGCTTGGGAGGAGGAGGCGCGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 937
 263 TyrGlyAlaProProAlaGlyTyrGlyAlaProProAlaGlyAlaLeuProPro 282
 938 TACCAAGACTGACATTTGTGAGCATTAAGGCGCGCGCTCTCGAGAG--GTGCGCGC 994
 282 alaTyrGlyAlaProSerAlaGlyAsnThrAlaAlaSerHisArgSerMetThr 300
 995 CAGCTTCCATGCCAGCGCGAGCCGACCAAGTGTGCGCGCTCTCTCTCTCTCTCTCT 1049

RESULT 10
 BC010616 1912 bp mRNA linear PRI 12-JUL-2001
 LOCUS Homo sapiens, clone MGC:18269 IMAGE:4177319, mRNA, complete cds.
 DEFINITION BC010616
 ACCESSION BC010616.1 GI:14714919
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS
 TITL
 JOURNAL

REMARK

COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: gcgaps-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
 Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: Villalobos, tmc.edu.
 Villalobos, D.K., Luna, R.A., Hale, S.M., Huylk, S., Lu, X., Garcia,
 A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
 Muzny, D.M., Gibbs, R.A.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
 Series: IRAT Plate: 12 Row: 9 Column: 23
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

FEATURES

SOURCE

1..1912
 /organism="Homo sapiens"

Alignment Scores:

Pred. No.: 9.32e-16 Length: 793
 Score: 460.00 Matches: 115
 Percent Similarity: 49.68% Conservative: 41
 Best Local Similarity: 36.62% Mismatches: 86
 Query Match: 26.92% Indels: 72
 DB: 10 Gaps: 12

US-09-864-291-5 (1-313) x AP499026 (1-793)

```

Oy 1 MetAlaValAenGlnSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 1 ATGGGCTCAACAAAGATCACTCAAGAGGC---GGCGAGTGAATGCTCAACACCTAG 57
Oy 21 SerValLeuLysGlnCysGluAspValAspLeuGlySerPheLeuGlnLysProValGluSer 40
Db 58 AGCATCTTAATGCTCAATGATCATGTGAATCACTTCACTCAACGACATGAAGAAATACCA 117
Oy 41 TyrLeuPheAenGlyThrLysLysGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
Db 118 GAGGCTTCAAGAGGACCAAGAAAGGACCGCTCACTCACTCACTCACTCACTCACTTT 177
Oy 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
Db 178 CTGTCCAAGGC-----AAGGATGCCATGCGCTCTTCAATGATGCCCTTCACTCACTG 231
Oy 81 SerAspCysThrIleGluGlnProIlePheAlaProAlaIleLysGlyThrIleGln 100
Db 232 AAGACTGTGAGTCAAGACCGCGGTGTGTGTCCTCACTTCAATGAAGGAGACATGAAA 291
Oy 101 AlaAlaProGlyGlyGlyTyrGluGlnGlnAlaValAlaPheLysLeuSerPheArgLysGly 120
Db 232 GCTGAAGACGAGAGGTGTGGAGAGGCTGCTCTTCACTCACTCACTCACTCACTCACTG 351
Oy 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaAlaArgGly--- 139
Db 352 GGTGCATTGAGTTGTGGGACGACAGATGCTCCAGGTGACATCTCAAGCTCCAGAGGTGA 411
Oy 140 IleProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrVal 159
Db 412 GTCCCAATGAGACCTATGAGTACCTTCACTGATGCCAGCGGGGCTCATGTCTTCCCA 471
Oy 160 ProGlyAlaAlaLysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
Db 472 CCA-----GTGCCAATGAGAAATGATACCTCTCCCTCGGCTCACTCACTCACTCACT 513
Oy 179 TyrGlyProProProProGlyTyrThrValGlnProGlyGlyTyrThrProProGlu 198
Db 514 TATCCACCGCGCCCGCCAGCGAGTTC----- 537
Oy 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProProMetGlyTyrGlyAlaProPro 218
Db 538 -----TACCTGCGGCTCTATATGATGACCGGGCC----- 567
Oy 219 ValGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValPro 238
Db 566 ATGGGATATGACAGCCCCA----- 588
Oy 239 ProGlyGlyTyrGlyAlaProProGlyGlyTyrGlyValProProGlyGlyTyrGlyAla 258
Db 589 -----CCACCA----- 594
Oy 259 ProProGlyGlyTyrGlyAlaProPro-----AlaGlyTyrGlyAlaProProAlaGly 276
Db 595 CCTTATCTGGGCGCCATGAGGCTCCCGTGTCAAGGGCCCAAGTGGCCCCCTTACT--- 651
Oy 277 AsnGluAlaLeuProProAlaTyrGluAlaProSerAlaGlyAsnThrAlaAlaSerHis 296
Db 652 -----CTGCACTGAGGCGCAAGGCTGCAAGACCGGCTGCCAGCGCTAT 696
Oy 297 ArgSerMetThrAlaGlnGlnGlnThrSerLeuProThrThr 310
Db 697 TACAACCGGAGCAACCAACAATGCTATCATGCCCCACGAGC 738
  
```

RESULT 12
 AR220846 1885 bp DNA linear PAT 26-SBP-2002
 LOCUS Sequence 87 from patent US 6426186.
 DEFINITION AR220846
 ACCESSION AR220846
 VERSION AR220846.1 GI:23327723
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE Unpublished.
 1 (bases 1 to 1885)
 AUTHORS Jones,K.A., Volkmuth,W. and Walker,M.G.
 TITLE Bone remodeling genes
 JOURNAL Patent: US 6426186-A 87 30-JUL-2002;
 FEATURES location/Qualifiers
 source 1..1885
 /organism="unknown"
 BASE COUNT 384 a 610 c 476 g 415 t
 ORIGIN

Alignment Scores:

Pred. No.: 2.26e-15 Length: 1885
 Score: 459.50 Matches: 129
 Percent Similarity: 45.43% Conservative: 40
 Best Local Similarity: 34.68% Mismatches: 100
 Query Match: 26.89% Indels: 104
 DB: 6 Gaps: 17

US-09-864-291-5 (1-313) x AR220846 (1-1885)

```

Oy 1 MetAlaValAenGlnSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 61 ATGGGCTCAACAAAGATCACTCGAGAGGC---GGCGAGTGAATGCTCAACACCTAG 117
Oy 21 SerValLeuLysGlnCysGluAspValAspLeuGlySerPheLeuGlnLysProValGluSer 40
Db 118 AGCATCTTAATGCTCAATGATCATGTGAATCACTTCACTCACTCACTCACTCACTT 177
Oy 41 TyrLeuPheAenGlyThrLysLysGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
Db 178 GAGGCTTCAAGAGGACCAAGAAAGGACCTGTCACTTCACTCACTCACTCACTCACTT 237
Oy 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaAlaArgGly--- 139
Db 352 GGTGCATTGAGTTGTGGAGAGGCTGCTCTTCACTCACTCACTTCACTTCACTCACTG 411
Oy 140 IleProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrVal 159
Db 472 GTCCCAATGAGACCTATGAGTACCTTCACTGATGCCAGCGGGGCTATGTCATGCCCG 531
Oy 160 ProGlyAlaAlaLysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
Db 532 CCA-----GTGCCAATGAGAAATGATACCTCTCGGCTCACTCACTCACTCACT 573
Oy 179 TyrGlyProProProPro-----Gly 185
Db 574 TATCCACCGCGCCCGCCAGGCTTCAATCCAGAACCCCGCATGAGACGGGCGCATGGA 633
Oy 186 TyrThrValGlnProGlyGlyTyrGlyThrProProGluGlyTyrGlyAlaGlnProGly 205
  
```

[illegible]

```

SOURCE
1..1802
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/dev_stage="16 day embryo"
40..825
/note="WBP-2; putative protein ligand to the WW domain of
yes-kinase associated protein YAP"
yes-kinase associated protein YAP
/db_xref="GI:177579"
/translation="MAALKRHSBGGVIVNTSILSYDHVELTFRDQVPEARPR
TKGVYTLPRIVITLSTKQAKMSFPMHPTLKDCKIPVRANITKIVAEKAG
GMSGSASYLRTFTAGALAEFGQRLQVASQASRKBVNGAYGVPMSSGAYVPP
NMTYPPPGYPPPPPPFYFGPPPMQAMGYVPPPPFYGMPPVSGSPSAPAT
AEKAAEAAASAYVNRGPHNVPTSQPPPPFYPPEDKKTY"

BASE COUNT      396 a      565 c      437 g      404 t
ORIGIN
Alignment Scores:
Pred. No.:      3..27e-15      Length:      1802
Score:          456.00      Matches:      123
Percent Similarity: 49.53%      Conservative: 35
Best Local Similarity: 38.56%      Mismatches:  94
Query Match:    26.68%      Indels:      67
DB:              10      Gaps:      14

US-09-864-291-5 (1-313) x MMU040826 (1-1802)
QY      1 MetAlaValAmGlnSerHisThrGluSerArgArgAlaLeuIleProSerGlyGlu 20
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      40 ATGGCGCTCAACAGAGATCACTCAAGAGGCG--GGCGAGAGATCGTCAACACAGTGA 96
QY      21 SerValLeuLysGlnGlyAspAlaIlePheLeuGlnGlyProValGluSer 40
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      97 AGCACTCTATAGTCTTATGATCATGTGGAGCTTACTTCAACACAGAGATGAGTCCCA 156
QY      41 TyrIlePheAmGlyThrIleValGlyThrIlePheLeuThrSerTyrArgValIlePhe 60
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      157 GAAGCGCTCAAGAGGACCAAGAAAGGCAACCGTCACTTACCTGACCGGATCTTT 216
QY      61 ValThrSerHisLeuValAmAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      217 CTGTCTTAAGGGG-----AAGGACCCCAATGACGTCTTCATGATGACCTTTTACCTGATG 270
QY      81 SerAspCysThrIleGlnIleProIlePheIleAlaProLeuTyrIleValGlyThrIleGln 100
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      271 AAGACCTGTAGATTAAAGACGCCGGTGTGTGCGAATTCATTAAAGGAGATGATGAG 330
QY      101 AlaIleAspProIleValIleTyrGlnGlyGlnAlaValPheLysLeuSerPheArgGlyGly 120
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      331 GCTGAAGCAGAGAGTGGCTGGGAAGGCTCCGCTCTTCAAGCTGACCTTACACAGAGG 390
QY      121 GlyAlaIleGlnPheAlaGlnLeuMetValIleAlaIleAspAlaIleAlaArgGly--- 139
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      391 GGCGCCATTGATGTTGGGACAGAGATGCTCCAGGTGCGATGCAAGCCTCCAGAGGGTGA 450
QY      140 IleProLeuGlnSerValLeuTyrThrPheAspThrSerGlyLeuTyrIleIleThrVal 155
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      451 GTCCCAATGAGACCTTATGAGTCAATGACCAAGCGGGGCTTATGCTTTCCTCCCG 510
QY      160 ProGlyValAlaValCysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      511 CCA-----GTCCCAATGAGATGTAACCTTCCTCCCTCTGGCTACCC----- 555
QY      179 TyrGlyProProProProGlyTyrThrValGlnProGlyIleTyrGlyThrProProGlu 198
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      553 TATCCACACACCCCAACCGAGTTC----- 576
QY      199 GlyTyrGlyValAlaGlnProGlyIleTyrValAlaProProMetGlyTyrGlyValAlaProPro 218
      |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db      577 -----TATCCAGACCTTCCATGATGATGATGGGGC----- 606

```

```

OY      219 ValGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValPro 238
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      607 ATGGGCTATGTACAGCCCA-----CCCA 630

OY      239 ProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValA 258
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      631 CCA-----CCCTATCTCGGGCCCAAGAGAGCTCGGTCAGAGCCCAAGTGC 678

OY      259 ProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyVal 278
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      679 CCC-----GCTACTCTGCA-----GCCAGGCCAAGGCTGCA 714

OY      279 AlaLeuProAlaTyrGlyValProSerAlaGlyValN----- 291
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      715 GCGCGTCGAGTGCCTATTAACACCA-----GGCAACCAACACATGCTACATGCC 768

OY      292 Thr-AlaAlaSerHisArgSerMetThrAlaGlnGlnGlnTherLeuProThr 309
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      769 ACGAGCAGCCTCCACCAACCACTACTACCCCGAGAGACAGAGACCAGT 823

RESULT 14
HS250D10
LOCUS   220895 bp  DNA  linear  PRI 05-JUN-2003
DEFINITION
Human DNA sequence from clone CTA-250D10 on chromosome 22. Contains
the genes for SREBF2 (sterol regulatory element binding
transcription factor 2), NAGA (alpha-N-acetylglucosaminidase), a
gene similar to neuronal-specific septin 3, a pseudogene similar to
ANP2 (adenine nucleotide translocator 2), 2 mRNAs based on ESTs, a
genomic marker D2S1178, a CA repeat polymorphism, ESTs and a CpG
island, complete sequence.
239716
ACCESSION
239716.4 GI:4456457
VERSION
HTG: ANP2; CpG Island; D2S1178; NAGA; septin 3; SREBF2.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 220895)
REFERENCE
AUTHORS
Clark G.
TITLES
Direct Submision
JOURNAL
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk
On Mar 21, 1999 this sequence version replaced gi:4164339.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
CTA-250D10 is
from the human BAC library described in U-J. Kim et al. (1996)
Genomics 34, 213-218.
VECTOR: pBAC108L
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.

```

FEATURES

```

source
1..220895
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="22"
/clone="CTA-250D10"
/clone_1lb="CT19788K-A1"
17..25
/note="2.6 copies 5 mer ATTG 26% conserved"
33..325
/note="AluX repeat: matches 1..301 of consensus"
326..338
/note="2.6 copies 5 mer TTGA 26% conserved"
667..683
/note="2.1 copies 8 mer AGAGACC 34% conserved"
740..751
/note="2.4 copies 5 mer TTTTA 24% conserved"
1443..1457
/note="7.5 copies 2 mer AC 21% conserved"
1715..1997
/note="AluDo repeat: matches 5..293 of consensus"
2103..2378
/note="AluBg repeat: matches 2..296 of consensus"
2844..2855
/note="12.0 copies 1 mer A 24% conserved"
complement(2884..3033)
/note="MIR repeat: matches 70..235 of consensus"
complement(3040..3344)
/note="AluY repeat: matches 1..310 of consensus"
3446..3681
/note="L1MB4 repeat: matches 5256..5471 of consensus"
3746..3837
/note="L1MB4 repeat: matches 5615..5707 of consensus"
3838..4142
/note="AluDo repeat: matches 2..305 of consensus"
4143..4441
/note="L1MB4 repeat: matches 5707..5999 of consensus"
4442..4747
/note="AluSg repeat: matches 1..292 of consensus"
4748..4858
/note="L1MB4 repeat: matches 5999..6110 of consensus"
4881..5172
/note="AluSx repeat: matches 1..296 of consensus"
5418..5429
/note="2.4 copies 5 mer GCTGA 24% conserved"
complement(5769..6021)
/note="match: GSS: Em:AQ069839"
6074..6089
/note="2.0 copies 8 mer CCCGGCTG 32% conserved"
6139..6153
/note="2.5 copies 6 mer TTTTA 21% conserved"
6141..6155
/note="3.0 copies 5 mer TTTAA 21% conserved"
complement(6366..6572)
/note="MIR repeat: matches 5..214 of consensus"
6839..6852
/note="2.3 copies 6 mer GGAAT 28% conserved"
7088..7114
/note="6.8 copies 4 mer TTTA 36% conserved"
7091..7109
/note="1.9 copies 10 mer ATTATTTAC 38% conserved"
complement(7122..7416)
/note="AluX repeat: matches 1..300 of consensus"

```

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>. This sequence is the entire insert of clone CTA-250D10. The true left end of clone RP1-18601 is at 129979 in this sequence. The true right end of clone RP5-359016 is at 1335 in this sequence. The true right end of clone RP5-921011 is at 23458 in this sequence. The true right end of clone RP3-359016 is at 118711 in this sequence.

```

repeat_region /note="19.5 copies 2 mer AT 78% conserved"
repeat_region 7502..7524 /note="11.5 copies 2 mer AC 46% conserved"
repeat_region 7523..7538 /note="2.0 copies 8 mer CAAATAT 32% conserved"
repeat_region 7535..7563 /note="14.5 copies 2 mer AT 49% conserved"
repeat_region /note="14.5 copies 2 mer AT 49% conserved"
repeat_region /note="14.5 copies 2 mer AT 49% conserved"
repeat_region 7873..7893 /note="5.2 copies 4 mer TGT 33% conserved"
repeat_region 7914..8200 /note="14.5 copies 2 mer AT 49% conserved"
repeat_region /note="14.5 copies 2 mer AT 49% conserved"
repeat_region 8387..8409 /note="11.5 copies 2 mer AC 46% conserved"
repeat_region 8483..8498 /note="2.0 copies 8 mer GTGGTGT 32% conserved"
repeat_region 8486..8496 /note="2.2 copies 5 mer GTGGT 22% conserved"
repeat_region 8606..8622 /note="2.4 copies 7 mer AGAGAA 34% conserved"
repeat_region /note="2.4 copies 7 mer AGAGAA 34% conserved"
repeat_region 9004..9014 /note="2.2 copies 5 mer CCTGG 22% conserved"
repeat_region 9037..9140 /note="MIR repeat: matches 17..123 of consensus"
repeat_region 9509..9823 /note="14.5 copies 2 mer AT 49% conserved"
repeat_region 9872..9883 /note="2.4 copies 5 mer GTCTT 24% conserved"
repeat_region 10071..49606 /note="2.4 copies 5 mer GTCTT 24% conserved"
gene /gene="SRBE2"
mRNA join<10071..10520,11851..12032,14129..14275,17038..17259,
18577..18651,18783..18964,20469..20661,21182..21363,
23956..24232,28082..28251,36357..36523,38060..38177,
40292..40401,41889..42021,43570..43738,46210..46395,
48103..48214,48680..49606)
/gene="SRBE2"
/product="bK250D10.1 (sterol regulatory element binding
transcription factor 2)"
/note="sterol regulatory element binding transcription
factor 2"
match: CDNA: Em:U02031 Em:U12330 Em:U0968 Em:U09103
Em:L16995 Em:U22818 Em:U22819 Em:U12322 Em:AB017337
match: ESTs: Em:R36312 Em:AL040138 Em:H92808 Em:AA053886"
/evidence=not experimental
join<10071..10520,11851..12032,14129..14275,17038..17259,
18577..18651,18783..18964,20469..20661,21182..21363,
23956..24232,28082..28251,36357..36523,38060..38177,
40292..40401,41889..42021,43570..43738,46210..46395,
48103..48214,48680..49606)
/gene="SRBE2"
/note="match: Proteins: Sw:Q12772 Sw:Q60429 Sw:P36956
Sw:Q60416 Tr:O15352"
/codon_start=3
/evidence=not experimental
/product="bK250D10.1 (sterol regulatory element binding
transcription factor 2)"
/protein_id="CAB41234.1"
/db_xref="GI:5578715"
/translation="MLQVSNQVGFPPDLPSFRLCSPPGSGSGSSSGSSSS
NGRSGSSGADVSVORSPTQVTLPSFSPSAAPQATQVKSPTSPPTPATILQ
PRPOPOPOBOLOOQVWITPTSTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
SSQVQPTVIOQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
KTDSLVTTLTKTDSVMAVAQNPALVTALTTTPTTPTTPTTPTTPTTPTTPTT
GOBKVPIKQVPGVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV
HKSQVLRKAIADYIKYLOQVNHKLQVNNWKLKANKLKDIDGSLVDNVDKIS
DENQVNLMSPADYDSGQAFSPYSIDBSPKPLDKVNDKVDGSLVDNVDKIS
RILLCVITFLCISFNPNTSLIQMGAHDSQHPHSGSRSVSFSSGSGMDMMPT

```

```

L1LMVNGVITLSPVFKLLVHGEPIVREHSHSVTPMHRRLQADLDLARGPFAAAGN
LQTCIAVARGALPTSRDLACSLRNVRVRLQDLRVRLTKVPOCRATPATBAG
FEDAKTSARDAALVRIHOLITGTCTPASASQSDYMLCAVLAACAKRIPST
LVRIHILTAAMELTKRCGRKGFPLASYFLSRQSLCGPESHAVPDSLRKLCPLQCKP

Alignment Scores:
Pred. No.: 5,87e-13 Length: 220895
Score: 451.00 Matches: 84
Percent Similarity: 63.64% Conserved: 7
Best Local Similarity: 58.74% Mismatches: 42
Query Match: 26.39% Indels: 10
DB: 9 Gaps: 2

US-09-864-291-5 (1-313) x H6250D10 (1-220895)
QY 174 TyrProIleValIleTyrGlyProProProGlyTyrThrValGlnProGlyGluTyr 193
DB 169999 TTCACGTAATATGCTATGAGGCCCACTTCAGAGA-----TAT 170037
QY 194 GlyThrProProGluGlyTyrGlyValGlnProGlyGlyTyrGlyValProProMetGly 213
DB 170038 GAGGCCCACTCCCGATACGAGGCCCACTTCAGAGATATGAGCCCACTTCAGAGA 170097
QY 214 TyrGlyAlaProProValGlyTyrGlyValProProGlyGlyTyrGlyValProProGly 233
DB 170098 AATGAGGCCCGCGCTGTGATACAGAGCCCTCCTGCGATATGAGCCCACTCTT 170157
QY 234 GlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProPro 253
DB 170158 GATACGAGGCCCACTTCAGATATGAGGCCCACTTCAGATATGAGGCCCACT 170217
QY 254 GlyGlyTyrGlyValProProGlyGlyTyrGlyValProProAlaGlyTyrGlyValPro 273
DB 170218 CTGATATATGAGAACCCCACTTCGATATGAGGCCCACTTCGATATGAGGCCCA 170277
QY 274 ProAlaGlyValGluValAlaLeuProProAlaTyrGlyValProProAlaGlyValAsnThrAla 293
DB 170278 CCGTGAAGAAATGAGAGGCCCGCTGCGGATACAGAGCTCACTTCGATACAGAGCC 170337
QY 294 AlaSerHisArgSerMetThrAlaGln-----GlnGluThrSerLeuProThrThr 310
DB 170338 AGGCTTCAGAAATATACAGAGCCCACTTCGAAACAGAGGCTTCTTCCTCCCTGCC 170397
QY 311 SerSerSer 313
DB 170398 TCCTCTCTCT 170406

RESULT 15
AB072784 2170 bp mRNA linear PRI 22-FEB-2003
LOCUS DEFINITION Macaca fascicularis testis cDNA clone:Qesa-20807, full insert
sequence.
ACCESSION AB072784
VERSION AB072784.1 GI:16041171
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Macaca fascicularis (craab-eating macaque).
ORGANISM Macaca fascicularis
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecoidea; Macaca.
REFERENCE 1 Osada,N., Hida,M., Kueuda,J., Tanuma,R., Hirata,M., Suto,Y.,
Hirai,M., Terao,K., Sugano,S. and Hashimoto,K.
Cynomolgus monkey testicular cDNAs for discovery of novel human
genes in the human genome sequence
BMC Genomics 3 (1), 36 (2002).
12498619
JOURNAL PUBMED 2 (bases 1 to 2170)
AUTHORS Hashimoto,K., Osada,N., Hida,M., Kueuda,J. and Sugano,S.
TITLE Direct Submision
REFERENCE 1 Hashimoto,K., Osada,N., Hida,M., Kueuda,J. and Sugano,S.
JOURNAL TITLE Submitted (09-OCT-2001) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources, 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan

```


COMMENT
Lab host: TOP10
(E-mail: khasshi@nih.gov.jp, URL: <http://www.nih.gov.jp/yoken/genebank/>,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 11:00:59 ; Search time 262.166 Seconds

(without alignment) 3222.858 Million cell updates/sec

Title: US-09-864-291-5

Perfect score: 1709

Sequence: 1 MAVVQSHTRSRGALIPSGE.....ASHRSMTAQQETSLPTSSS 313

Scoring table:

BLOSUM62	Xgapop 10.0 , Ygapext 0.5
	Ygapop 10.0 , Ygapext 0.5
	Ygapop 6.0 , Ygapext 7.0
	Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODL=frame+g2n.model -DEV=x1p
-O=/cgn2_1/USPRO_epool_p/US09864291/runat_15122003_160858_23296/app_query.fasta_1.846
-DB=N Geneseq_19Jun03 -QWTF=fastcap -SUPFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LISTEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum2 -TRANS=human40.cdd
-LIST=45 -DOCALLIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODL=LOCAL -OUTFMT=pco -NOR=ext -HEAPSIZE=500 -MITLEN=0 -MITLEN=2000000000
-USER=US09864291.QCGN_1_1_511/runat_15122003_160858_23296 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOC
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2003.DAT.*
25: /SIDSI/gcgdata/geneseq/geneeqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1709	100.0	1413	24	AA520601
2	769	45.0	1001	24	AA520592
3	543	31.8	467	22	AAU01229
4	543	31.8	467	23	ABU96688
5	475	27.8	1915	21	AA223633
6	475	27.8	1915	25	AB273614
7	475	27.8	1915	25	AB273614
8	475	27.8	1915	25	AB273614
9	459.5	26.9	1885	24	AB570430
10	451	26.4	220895	24	ABX84798
11	438	25.6	436	22	ABA67850
12	438	25.6	436	22	AAK42003
13	438	25.6	436	22	AAI48070
14	438	25.6	436	22	AB516034
15	429.5	25.1	7099	22	AAU04882
16	429.5	25.1	7099	23	ABU97776
17	408	23.9	1109	23	ABU17151
18	347	20.3	321	22	AAU01355
19	347	20.3	321	22	ABU96808
20	321	18.8	454	21	AAU04014
21	320.5	18.8	409	25	ABX41819
22	310	18.1	4323	23	ABU17150
23	231	13.5	2614	23	ABU28430
24	231	13.5	4954	23	ABU05187
25	231	13.5	9516	23	ABU05186
26	228.5	13.4	894	23	AA570582
27	227	13.3	1863	23	ABU28035
28	227	13.3	4329	23	ABU28034
29	221.5	13.0	5751	23	ABU28034
30	219.5	12.8	3206	23	ABU16637
31	219.5	12.8	5595	23	ABU16636
32	218.5	12.8	1174	21	AAU06029
33	216	12.6	2016	24	AAU06029
34	215	12.6	1015	24	AAU06029
35	215	12.6	1812	21	AAU06029
36	214	12.5	1774	11	AAU06029
37	214	12.5	1774	24	ABK83773
38	214	12.5	2176	24	ABY94224
39	213.5	12.5	3222	23	ABU09835
40	211	12.3	758	22	AAU23504
41	210.5	12.3	4403765	22	AAU23504
42	210.5	12.3	4411529	22	AAU23504
43	209	12.2	758	22	AAU23504
44	208.5	12.2	623	23	ABU12315
45	208.5	12.2	2623	23	ABU12314

ALIGNMENTS

RESULT 1

AA520601 standard; cDNA, 1413 BP.

AA520601

09-APR-2002 (first entry)

DNA encoding bovine perinuclear theca 32 (PT32).

Testicular WM domain binding protein; htmpp; perinuclear theca 32;

PT32; contraceptive; fertility; oocyte activation; vaccine;

globozoospermia; spermatogenesis; spermatozoa; tyrosine kinase; c-Yes;

Immunonutrient; bovine; gene; ss.

XX

XX

XX

XX DNA encoding human testicular WW domain binding protein (htwbp).
 XX
 XX Testicular WW domain binding protein, WBP; perinuclear theca 32;
 XX PT32; contraceptive; fertility; oocyte activation; vaccine;
 XX globozoospermy; spermatogenesis; spermatozoa; tyrosine kinase; c-Yes;
 XX immunoreceptive; human; gene; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT primer_bind 1..18
 FT /tag= a
 FT /note= "Binds forward primer AAS20605"
 FT CDS 1..708
 FT /tag= b
 FT /product= "htwbp"
 FT /note= "Human testicular WW domain binding protein"
 FT primer_bind complement (984..1001)
 FT /tag= c
 FT /note= "Binds reverse primer AAS20606"
 XX
 XX WO200190185-A2.
 XX
 XX PD 29-NOV-2001.
 XX
 XX PD 25-MAY-2001; 2001WO-CA00738.
 XX
 XX PR 25-MAY-2000; 2000CA-2307128.
 XX PR 25-MAY-2000; 2000US-206979P.
 XX
 XX PA (TOOH) UNIV QUEENS KINGSTON.
 XX PA (UYOR-) UNIV OREGON HEALTH SCI.
 XX
 XX PI Olo R, Sutovsky P;
 XX
 XX DR MPI, 2002-097644/13.
 XX DR P-PSDB; AAU74610.
 XX
 XX PT Isolated perinuclear theca 32 polypeptide that interacts with activated
 XX tyrosine kinase c-Yes, for enhancing fertility, treating/diagnosing
 XX diminished fertility and abnormal spermatogenesis and for providing
 XX contraception -
 XX
 XX ES Claim 62; Fig 4B; 103pp; English.
 XX
 XX CC The invention describes an isolated perinuclear theca 32 (PT32)
 XX polypeptide (I) which interacts with tyrosine kinase c-Yes. (I) is
 XX useful for: enhancing fertility in a mammal; treating globozoospermy, by
 XX expressing (I) in spermatozoa; inhibiting fertilization, by introducing
 XX (I) or its antigenic fragment into a mammal to elicit an immune
 XX response; enhancing the ability of round spermatids to activate oocytes;
 XX treating or diagnosing diminished fertility and abnormal spermatogenesis;
 XX in providing contraception; identifying contraceptive and
 XX fertility-enhancing agents. The polynucleotide is useful for producing
 XX (I) by recombinant techniques, as vaccine, as diagnostic reagents, and
 XX for chromosome identification. An antibody against (I) is useful in
 XX immunological assays, in immunoreceptive methods, to identify cells
 XX expressing (I), and to purify (I) by affinity chromatography. A
 XX transgenic animal is useful as an animal model for studying human
 XX fertility and reproductive biology, and for screening compounds to
 XX identify modulators of oocyte activation. The use of (I) prevents the
 XX entry of components which are detrimental to embryonic development into
 XX the oocyte during oocyte activation with crude sperm extract and avoids
 XX the propagation of viruses such as HIV (human immunodeficiency virus) and
 XX SIV (simian immunodeficiency virus) carried in the sperm. This sequence
 XX encodes the human testicular WW domain binding protein (htwbp), described
 XX in the method of the invention.
 XX
 XX SQ Sequence 1001 BP; 261 A; 254 C; 248 G; 238 T; 0 other;
 Alignment Scores: 1.4e-34 Length: 1001
 Pred. No.:

Score: 769.00 Matches: 148
 Percent Similarity: 68.60% Conserved: 18
 Best Local Similarity: 61.16% Mismatches: 62
 Query Match: 45.00% Indels: 14
 DB: 24 Gaps: 4
 US-09-864-291-5 (1-313) x AAS20602 (1-1001)
 QY 75 MetProPheGlyLeuMetSerAspCysThrIleGluGlnProIlePheAlaProAsnTyr 94
 DB 1 ATGCCATTGATCTGATGACGACACCTCACTGTAAACACGATTTGCTGCAAACTTC 60
 QY 95 IleYsglyThrIleGlnAlaAlaProGlyGlyGlyTyrGlyGlyGlnAlaValPheYys 114
 DB 61 ATTAAAGGAACTTATTCAGGACGCTCCATATGATGCTGGAGAGACAACTTCTTAA 120
 QY 115 LeuSerPheArglyGlyGlyAlaIleGluPheAlaGlnLeuMetValYsAlaAser 134
 DB 121 TTAGCTTCAGAAATGAGAGATGCCATTGAATTTCCAGTTGATGGTGAAGCTGCTT 180
 QY 135 AlAlAlAlaArglyIleProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeu 154
 DB 181 GCTGTGCCCGAGGATTTCCACTTGAACCTTAATATGACTGCTTCACTGATGGAATT 240
 QY 155 TyrIleIleThrValProGlyAlaAlaValCysSerSerGlnThrProCysProAlaTyr 174
 DB 241 TATGTAATTAAGTGGGAAAGG--AATATGCACTCCACAGATCCCTTGT----- 288
 QY 175 ProIleValIleTyrGlyProProProProGlyTyrThrValGlnProGlyGluTyrGly 194
 DB 289 TCAGTATATGCTATGAGGAGCCCACTGACAG-----TATGGA 327
 QY 195 ThrProProGlyGlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProPrometGlyTyr 214
 DB 328 GCCCACTCCCGATACGAGAGGCCCACTGACAGATATGAGCCCAACCGTGAAT 387
 QY 215 GlyAlaProProValGlyTyrGlyValAlaProProGlyGlyTyrGlyValProProGly 234
 DB 388 GAAAGCCCGCTGTGGATACAGAGCTCACTGTGAGATATGAGCCCACTTTGGA 447
 QY 235 TyrGlyValProProGlyGlyTyrGlyValAlaProProGlyGlyTyrGlyValProProGly 254
 DB 448 TACGAGCCCACTGACAGATATGAGAGCCCACTGAGATATGAGCCCACTTT 507
 QY 255 GlyTyrGlyAlaProProGlyGlyTyrGlyValAlaProProAlaGlyTyrGlyAlaProPro 274
 DB 508 GGAATATGAAACCCCACTCTCGATATGAGCCCACTCTCGATATGAGCCCACT 567
 QY 275 AlaGlyGlnGluAlaLeuProProAlaTyrGlyAlaProSerAlaGlyAsnThrAlaAla 294
 DB 568 GCAGAAATGAAAGCCCGCTGCGGATACAGAGCTCACTGTGATCAGAGCAGG 627
 QY 295 SerIAsrSerMetThrAlaGln-----GlnIuThrSerLeuProThrThrSer 311
 DB 628 CTTAGAAATCTACAGACGCCAGGCTCTGAAAACAGAGGCTTCTTCCCTGCTCC 687
 QY 312 SerSer 313
 DB 688 TCTTCT 693
 RESULT 3
 AAL01229 standard; cDNA; 467 BP.
 ID AAL01229;
 AC AAL01229;
 XX 21-NOV-2001 (first entry)
 XX Human reproductive system related antigen cDNA SEQ ID NO: 1230.
 XX Human; reproductive system related antigen; reproductive system disorder;
 XX cancer; gene therapy; ss.
 XX


```
XX PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI, 2001-465570/50.
XX P-PSDB; AA095259.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 1; SEQ ID NO 1230; 1297pp + Sequence Listing, English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a coding sequence of the
XX invention.
XX
XX SQ Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,38e-22 Length: 467
XX Score: 543.00 Matches: 104
XX Percent Similarity: 69.46% Conservative: 12
XX Best Local Similarity: 62.28% Mismatches: 38
XX Query Match: 31.77% Indels: 13
XX DB: 22 Gaps: 4
XX
XX US-09-864-291-5 (1-313) x AA01229 (1-467)
XX
XX 118 ArglyGlyGlyValAlaIleGluPheAlaGlnLeuMetValIleAlaIleAlaAla 137
XX |||||
XX 3 AAAAAATGAGAGTGCCTTGAATTGCGCAAGTTGATGATGAAGCTCCCTGCTGCTGCC 62
XX
XX 138 ArgGlyIleProLeuGlySerValaAntYrTrpPheAspThrSerGlyLeuYrIleIle 157
XX |||||
XX 63 CAGAGATTCCACTTGAACCTTAATGACTGGTTCAGCTCATGGAATTTATGTAATT 122
XX
XX 158 ThrValProGlyValAlaIleValGlySerSerGlnThrProGlyProAlaIleYrProIleVal 177
XX |||||
XX 123 ACTGGGGAGAGG---AAATATGTGACTCCACAGATGCTTGT-----TCAGTTATT 170
XX
XX 178 IleYrGlyProProProProGlyYrThrValGlnProGlyGluYrGlyThrProPro 197
XX |||||
XX 171 GTCATAGAGCCCACTGACAGA-----TATGAGCCCACT 209
XX
XX 198 GluGlyYrGlyValAlaIleProGlyYrGlyValaProProMetGlyYrGlyValaPro 217
XX |||||
XX 210 CCGGATACGAGAGCCCACTGACAGATATGAGCCCAACCGTAGAATGAGAGCCCG 269
XX
XX 218 ProValGlyYrGlyValaProProGlyYrGlyValaProProGlyYrGlyVala 237
XX |||||
XX 270 CCTGTGGATACAGAGCCCTCAGTGTGATGAGCCCACTCTTGATACGAGGCC 329
XX
XX 238 ProProGlyYrGlyValaProProGlyYrGlyValaProProGlyYrGlyVala 257
XX |||||
XX 330 CCACTGCGAGATATGAGCCCACTCTGAGATATGAGCCCACTTTGAGATATG 389
XX
XX 257 ValaProProGlyYrGlyValaProProAlaGlyYrGlyValaProProAlaGlyYr 277
XX |||||
XX 390 AACCCCACTCTGAGATATGAGCCCACTTTCGATATGAGCCCACTGCGAGAAA 449
XX
XX 277 nGluAlaLeuProProAla 283
XX |||||
XX 450 TGAAGGC---CCGCTTCG 465
XX
XX RESULT 4
XX ABL96688
XX ID ABL96688 Standard; cDNA; 467 BP.
XX
XX ABL96688;
XX
XX 21-JUN-2002 (first entry)
```

```
XX
XX Human testicular antigen encoding cDNA SEQ ID NO: 356.
XX
XX Human; testicular antigen; testes; cancer; metastasis; immune disorder;
XX reproductive system disorder; urinary system disorder; gene therapy;
XX cardiovascular disorder; respiratory disorder; neurological disorder;
XX gastrointestinal disease; infection; cytostatic; gene; ss.
XX
XX Homo sapiens.
XX
XX MO200155317-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001MO-US01329.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
XX 07-JUL-2000; 2000US-0216880.
XX
XX 11-JUL-2000; 2000US-0217487.
XX
XX 11-JUL-2000; 2000US-0217496.
XX
XX 14-JUL-2000; 2000US-0218290.
XX
XX 26-JUL-2000; 2000US-0220963.
XX
XX 26-JUL-2000; 2000US-0220964.
XX
XX 14-AUG-2000; 2000US-0224518.
XX
XX 14-AUG-2000; 2000US-0224519.
XX
XX 14-AUG-2000; 2000US-0225213.
XX
XX 14-AUG-2000; 2000US-0225214.
XX
XX 14-AUG-2000; 2000US-0225266.
XX
XX 14-AUG-2000; 2000US-0225267.
XX
XX 14-AUG-2000; 2000US-0225268.
XX
XX 14-AUG-2000; 2000US-0225270.
XX
XX 14-AUG-2000; 2000US-0225447.
XX
XX 14-AUG-2000; 2000US-0225757.
XX
XX 14-AUG-2000; 2000US-0225758.
XX
XX 14-AUG-2000; 2000US-0225759.
XX
XX 18-AUG-2000; 2000US-0226275.
XX
XX 22-AUG-2000; 2000US-0226681.
XX
XX 22-AUG-2000; 2000US-0226868.
XX
XX 22-AUG-2000; 2000US-0227182.
XX
XX 23-AUG-2000; 2000US-0227009.
XX
XX 30-AUG-2000; 2000US-0228924.
XX
XX 01-SEP-2000; 2000US-0229287.
XX
XX 01-SEP-2000; 2000US-0229343.
XX
XX 01-SEP-2000; 2000US-0229344.
XX
XX 01-SEP-2000; 2000US-0229345.
XX
XX 05-SEP-2000; 2000US-0229509.
XX
XX 05-SEP-2000; 2000US-0229513.
XX
XX 06-SEP-2000; 2000US-0230437.
XX
XX 06-SEP-2000; 2000US-0230438.
XX
XX 08-SEP-2000; 2000US-0231242.
XX
XX 08-SEP-2000; 2000US-0231243.
XX
XX 08-SEP-2000; 2000US-0231244.
XX
XX 08-SEP-2000; 2000US-0231413.
XX
XX 08-SEP-2000; 2000US-0231414.
XX
XX 08-SEP-2000; 2000US-0232080.
XX
XX 08-SEP-2000; 2000US-0232081.
XX
XX 12-SEP-2000; 2000US-0231968.
XX
XX 14-SEP-2000; 2000US-0232397.
XX
XX 14-SEP-2000; 2000US-0232398.
XX
XX 14-SEP-2000; 2000US-0232399.
XX
XX 14-SEP-2000; 2000US-0232400.
XX
XX 14-SEP-2000; 2000US-0232401.
XX
XX 14-SEP-2000; 2000US-0233063.
```

PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234323
PR	21-SEP-2000	2000US-0234774
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0235484
PR	26-SEP-2000	2000US-0235498
PR	27-SEP-2000	2000US-0235634
PR	27-SEP-2000	2000US-0235636
PR	29-SEP-2000	2000US-0236527
PR	29-SEP-2000	2000US-0236567
PR	29-SEP-2000	2000US-0236568
PR	29-SEP-2000	2000US-0236569
PR	29-SEP-2000	2000US-0236570
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239935
PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241121
PR	20-OCT-2000	2000US-0241185
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0246177
PR	20-OCT-2000	2000US-0246178
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249298
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249310
PR	17-NOV-2000	2000US-0249311
PR	17-NOV-2000	2000US-0249312
PR	17-NOV-2000	2000US-0249313
PR	17-NOV-2000	2000US-0249314
PR	17-NOV-2000	2000US-0249315
PR	17-NOV-2000	2000US-0249316
PR	17-NOV-2000	2000US-0249317
PR	17-NOV-2000	2000US-0249318
PR	17-NOV-2000	2000US-0249319
PR	17-NOV-2000	2000US-0249320
PR	17-NOV-2000	2000US-0249321
PR	17-NOV-2000	2000US-0249322
PR	17-NOV-2000	2000US-0249323
PR	17-NOV-2000	2000US-0249324
PR	17-NOV-2000	2000US-0249325
PR	17-NOV-2000	2000US-0249326
PR	17-NOV-2000	2000US-0249327
PR	17-NOV-2000	2000US-0249328
PR	17-NOV-2000	2000US-0249329
PR	17-NOV-2000	2000US-0249330
PR	17-NOV-2000	2000US-0249331
PR	17-NOV-2000	2000US-0249332
PR	17-NOV-2000	2000US-0249333
PR	17-NOV-2000	2000US-0249334
PR	17-NOV-2000	2000US-0249335
PR	17-NOV-2000	2000US-0249336
PR	17-NOV-2000	2000US-0249337
PR	17-NOV-2000	2000US-0249338
PR	17-NOV-2000	2000US-0249339
PR	17-NOV-2000	2000US-0249340
PR	17-NOV-2000	2000US-0249341
PR	17-NOV-2000	2000US-0249342
PR	17-NOV-2000	2000US-0249343
PR	17-NOV-2000	2000US-0249344
PR	17-NOV-2000	2000US-0249345
PR	17-NOV-2000	2000US-0249346
PR	17-NOV-2000	2000US-0249347
PR	17-NOV-2000	2000US-0249348
PR	17-NOV-2000	2000US-0249349
PR	17-NOV-2000	2000US-0249350
PR	17-NOV-2000	2000US-0249351
PR	17-NOV-2000	2000US-0249352
PR	17-NOV-2000	2000US-0249353
PR	17-NOV-2000	2000US-0249354
PR	17-NOV-2000	2000US-0249355
PR	17-NOV-2000	2000US-0249356
PR	17-NOV-2000	2000US-0249357
PR	17-NOV-2000	2000US-0249358
PR	17-NOV-2000	2000US-0249359
PR	17-NOV-2000	2000US-0249360
PR	17-NOV-2000	2000US-0249361
PR	17-NOV-2000	2000US-0249362

	PR	08-DEC-2000;	2000US-0251869.	
	PR	08-DGC-2000;	2000US-0251989.	
	PR	08-DRC-2000;	2000US-0251990.	
	PR	11-DEC-2000;	2000US-0254097.	
	PR	05-JAN-2001;	2001US-0253678.	
XX		(HUMA-) HUMAN GENOME SCI INC.		
PA		Rosen CA, Barash SC, Ruben SM;		
PI		WPI; 2001-483232/52.		
DR		Nucleic acids encoding 973 human testicular antigen polypeptides,		
PT		useful for preventing, diagnosing and/or treating testicular cancer -		
PS		Claim 1, SEQ ID NO 356; 766pp; English.		
XX		The present invention provides the protein and coding sequences of 973		
CC		human testicular antigens, and fragments of their genomic sequences. The		
CC		sequences can be used in the treatment of cardiovascular, urinary system,		
CC		reproductive system, immune, respiratory, neurological and		
CC		gastrointestinal disorders, infections, and particularly cancer,		
CC		especially testicular cancers. The present sequence is a cDNA of the		
CC		Invention.		
XX				
SO		Sequence 467 BP; 109 A; 132 C; 124 G; 100 T; 2 other:		
	Alignment Scores:			
	Pred. No.:	2.38e-22	Length:	467
	Score:	543.00	Matches:	104
	Percent Similarity:	69.46%	Conservative:	12
	Best Local Similarity:	62.28%	Mismatches:	38
	Query Match:	31.77%	Indels:	13
	DB:	23	Gaps:	4
US-09-864-291-5 (1-313) X ABL96688 (1-467)				
OY		118 ArgLYVGILVGLIALIIGLIuPhEALAGLLeuECtVALyBALalAsERAlAlAlA		1377
Dd		3 AGAAATGAGGGGTGCATTGAATTTGCCAGTGTGATGTGAAGCTGCTCGCTGCC		62
OY		138 ArgGLYLIPRoLeucljSeRvAlAsnTyTrPheApThRSerGlYleuTYlAlle		157
Dd		63 CGAGGATTCACCTTAAGAACCCTTAATAACTGTCAGCTTATGGAAATTATGTAATT		122
OY		158 ThrValProGLyAlAlAlAVAlCySeRSerGIInThrProCYSProAlArzProIIeVal		177
Dd		123 ACTGGGGAAAGG---AAATATGCACTCACAAGATGCTT-----TGAGTTATT		170
OY		178 ILleTYrgLYPrOPROPrOBrogLYTYThrValGlnPROglYgluTYrgLYThrPro		197
Dd		171 GTCATAGAGAGCCCACTGCAGAG-----TAGAGCCCCCACT		208
OY		198 GLUGLYTYRGlyVALAGLnPROglYgluTYrgLYVALAPROPromEGLYTYrgLYAlAPRO		217
Dd		210 CCCGATATGAGAGGCCCACTGCAGAGATATGAGGCCAACCCGTAGAAATATGAAAGCCG		265
OY		218 ProValGLYTyrGLyVALAPROPrOGlygluTYrgLYVALProPROglYgluTYrgLYVAL		237
Dd		270 CCTGTGGATACAGAGCCTCACCTGTGCATATGAGGCCCACTCTTGATATCAGAGCC		32
OY		238 PROPrOGlygluTYrgLYVALAPROPrOGlygluTYrgLYVALPro-ProGLYgluTYrgLY		257
Dd		330 CCACCTGCAGAGATATGAGGCCCACTCTAGAGATATGAGGCCCACTTGTGATATG		388
OY		257 yAlAPROPrOGlygluTYrgLYVALAPROPrOAlAGLyTYrgLYVALAPROPrOAlAGLyAS		277
Dd		390 AACCCCACTCTCGATATGAGGCCCACTTGTGATATGAGGCCCACTTCAGAGAAA		449
OY		277 ngluAlAleuPrOPROAlA		283
Dd		450 TGAAAGC---CCGCTTTCG		465

RESULT 5
 AAF22363
 ID AAF22363 standard; cDNA; 1915 BP.
 AC AAF22363;
 XX
 XX 26-MAR-2001 (first entry)
 XX
 XX Human secreted protein gene 48 SEQ ID NO:58.
 XX
 XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 XX cerebroprotective; neuroprotective; antibacterial; virucide;
 XX fungicide; ophthalmological; vulnary; gene therapy; neoplasm;
 XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 XX cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 XX cerebral ischemia; angiogenesis; nervous system disorder; infection;
 XX Alzheimer's disease; ocular disorder; corneal infection; wound healing;
 XX skin aging; food additive; preservative; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO20061748-A1.
 XX
 XX 19-OCT-2000.
 XX
 XX 06-APR-2000; 2000WO-US08982.
 XX
 XX 09-APR-1999; 99US-0128696.
 XX
 XX 14-JAN-2000; 2000US-0176069.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruden SM, Komatsu G;
 XX
 DR WPI; 2000-638566/61.
 DR
 XX P-PSDB; AAB63096.
 XX
 XX New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PT
 XX
 PS Claim 1; Page 429-430; 480pp; English.
 XX
 CC AAF22363 encode the human secreted proteins given in AAB63049
 CC to AAB63096. AAB63097 to AAB63132 represent more human secreted proteins
 CC and polypeptides homologous to them. Human secreted proteins have
 CC activities based on the tissues and cells the genes are expressed in.
 CC Examples of activities include: immunosuppressive; antiarthritic;
 CC antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 CC cerebroprotective; neuroprotective; antibacterial; virucide;
 CC fungicide; ophthalmological; and vulnary. The polynucleotides and
 CC proteins can be used to prevent, treat or ameliorate a medical condition
 CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
 CC sheep. They are also used in diagnosing a pathological condition or
 CC susceptibility to a pathological condition. Disorders which are diagnosed
 CC or treated include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi and
 CC ocular disorders e.g. corneal infection. The polypeptides can also be
 CC used to aid wound healing and epithelial cell proliferation, to prevent
 CC skin aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis. The polypeptides can also be used as a food additive or
 CC preservative to increase or decrease storage capabilities, fat content,
 CC lipid, protein, carbohydrate, vitamins, minerals, cofactors and other
 CC nutritional components. AAF22307 to AAF22315 and AAB63048 represent
 CC sequences used in the exemplification of the present invention.
 CC
 XX Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;

Alignment Scores:
 Pred. No.: 5,15e-18 Length: 1915
 Score: 475.00 Matches: 132
 Percent Similarity: 47.08% Conservative: 37
 Best Local Similarity: 36.77% Mismatch: 113
 Query Match: 27.79% Indels: 77
 DB: 21 Gaps: 16
 US-09-864-291-5 (1-313) x AAF22363 (1-1915)
 QY 1 MetAlaValAngInSerHisThrGluSerArgArgAlaLeuIleProSerGlyGlu 20
 DB 59 ATGGCGCTCAACAAGATCACTCGAGGCGC---GGCGAGATGATGTCATATACACCGAG 115
 QY 21 SerValLeuGlyGlnGlyGluAspValAspLeuGlySerPheLeuGlnLysProValGluSer 40
 DB 116 AGCATCTTAATGCTTAATGCTTAATGCTTAATGCTTAATGCTTAATGCTTAATGCTTAAT 175
 QY 41 TyrLeuPheAngInGlyThrLysGlyGlyThrLeuPheLeuThrSerTyrArgValPhe 60
 DB 176 GAAGCTTCAAG 235
 QY 61 ValThrSerHisLeuValAspAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 236 CTGTCAAGGCGC-----AAGATGCGCATGATGCTTCAATGATGCTTATATCTCATG 289
 QY 81 SerAspGlyThrIleGluGlnProIlePheAlaProMetTyrIleGlyGlyThrIleGln 100
 DB 290 AAAGCTGTGAGATCAACAGACCGCTATTGTGTCAATCAATCAATCAATCAATCAATCAAT 349
 QY 101 AlaAlaProGlyGlyGlyTyrGlyGlyGlyAlaValPheLysLeuSerPheArgLysGly 120
 DB 350 GCGGAGCGGAG 409
 QY 121 GlyAlaIleGluPheAlaGlnLeuMetValIleAlaValPheLysLeuSerPheArgLysGly 139
 DB 410 GCGCGCATGTGATTCGACAGACGATGCTCAAGGATGATCAAGGATGATCAAGGATGATCAAG 469
 QY 140 IleProLeuGlySerValAspTyrTyrPheAspThrSerGlyLysLeuTyrIleThrVal 159
 DB 470 GTCCCGCATGTGAG 529
 QY 160 ProGlyAlaAlaValCysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
 DB 530 CCA-----GTCCCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 571
 QY 179 TyrGlyProProProProGlyTyrThrValGlnProGlyGlyGlyTyrThrProProGly 198
 DB 572 TATCAACCGCCCGCCACCTGAGTTC-----TATCAAGACCCCGCAG 613
 QY 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProPro-----MetGlyTyr 214
 DB 614 ATGACGCGGCGCAG---GATATGATGACGCGCCACACGCGCCTACCTGAGGCC 667
 QY 215 GlyAlaProProVal---GlyTyrGlyAlaPro----- 224
 DB 668 ATGAACTCTCGTCAAGGCGCCGATGTCCTTCACTCTGACGCGAAGCGAGGCC 727
 QY 225 ----- 727
 DB 728 GCAAGAGAGCGCGCAGCGCTATTACACCCAGAGCAATCTTCAACAGTCTTACAGTCC 787
 QY 226 GlyGlyTyrGlyValProProGlyGlyTyrGly----- 236
 DB 788 ACGAGCGAGCGCGCGCCTCTTCACTCAACCGAGATGAAGAACCGAGTAGGCC 847
 QY 237 -----ValProProGlyGly---TyrGlyAlaProPro----- 246
 DB 848 CTCCTGCTCTCTGCTCTCTCACTCTCACTCTCACTCTCACTCTCACTCTCACTCTCTG 907
 QY 247 GlyGlyTyrGlyValProProGlyGlyTyrGlyAlaProProGly-----Gly 262
 DB 908 CTGGGCTTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 967

QY 263 TyrGlyAlaProPro-AlaGlyTyrGlyAlaProProAlaGlyAlaSerGlyAlaLeuProP- 282
 DB 968 TACAGAGACTACATTCATTCGACATTCAGGCCCCCGCTCCGAGAG---GTGCGCC 1024
 QY 282 caATyrGlyAlaProSerAlaGlyAlaSerAlaAlaSerAlaArgSerMetThr 300
 DB 1025 CAGCTTCATTCGACGCGGAGCCGACAGTGTGCGCCAGCGTACCTCCCTCACC 1079
 RESULT 6
 AB273614
 ID AB273614 standard; cDNA; 1915 BP.
 AC AB273614;
 XX
 DT 12-MAY-2003 (first entry)
 XX
 DE Secreted protein-encoding gene 334 cDNA clone HTLEML6, SEQ ID NO:344.
 XX
 KW Human; secreted protein; cancer; tumour; hyperproliferative disorder;
 KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
 KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
 KW drug screening; chromosome identification; chromosome mapping;
 KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
 KW antihaemic; vulnery; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200277013-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002MO-US09370.
 XX
 PR 27-MAR-2001; 2001US-278650P.
 PR 12-SEP-2001; 2001US-0950082.
 PR 12-SEP-2001; 2001US-0950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2003-040578/03.
 DR P-PSDB; ABR01280.
 XX
 PT New human secreted proteins and nucleic acids, useful for detecting or
 PT treating cancer or other hyperproliferative disorders, autoimmune
 PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia -
 XX
 PS Claim 21; Page 1332; 2474pp; English.
 XX
 AB273281-AB273697 represent cDNAs corresponding to 391 human secreted
 CC protein genes, and ABP00947-ABP01363 represent the proteins they encode.
 CC AB273698-AB274687 represent human secreted protein genomic fragments. The
 CC invention also encompasses antibodies specific for the secreted proteins,
 CC the use of the secreted proteins in drug screening and recombinant
 CC vectors and host cells comprising a nucleic acid of the invention. The
 CC secreted proteins are thought to be involved in biological activities
 CC associated with cellular signalling, cellular differentiation, cell
 CC migration, prohormone activation and neurotransmitter activity. The
 CC secreted proteins, nucleic acids encoding them, antibodies or antibody
 CC fragments specific for the secreted proteins, and modulators of protein
 CC activity are useful for diagnosing or treating cancers or other
 CC hyperproliferative disorders. Additionally, the secreted proteins and
 CC their nucleic acids may also be used in the treatment of autoimmune
 CC disorder, inflammatory disorders, diseases involving angiogenesis, AIDS
 CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
 CC wound healing. Nucleic acids of the invention may be used for chromosome
 CC identification, chromosome mapping, in gene therapy, for identifying
 CC individuals from minute biological samples, as hybridisation probes, and
 CC as molecular weight markers. The present sequence represents a human
 CC secreted protein-encoding cDNA clone of the invention.

BQ Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;
 Alignment Scores:
 Pred. No.: 5, 15e-18 Length: 1915
 Score: 475.00 Matches: 132
 Percent Similarity: 47.08% Conserved: 37
 Best Local Similarity: 36.77% Mismatches: 113
 Query Match: 27.79% Indels: 77
 DB: 25 Gaps: 16
 US-09-864-291-5 (1-313) x AB273614 (1-1915)
 QY 1 MetAlaValAsnGlnSerHisThrGluSerArgArgGlyAlaLeu1LeuProSerGlyGlu 20
 DB 59 ATGGCGCTCAACAAGATCATCTCGAGGCGC---GGCGAGATGATGTGTAATACACCGAG 115
 QY 21 SerValLeuLysGlnCysGluArgValAlaLeuCysPheLeuGlnLysProValGluSer 40
 DB 116 AGCATCTTAATGTCTTATGATCACTGGAACCTCAATTCATGATGATGAGAACGTGCCA 175
 QY 41 TyrLeuPheAsnGlyThrLysLysGlyThrLeuPheLeuThrSerCysArgValAlaPhe 60
 DB 176 GAAGCTTCAAAAGGACCAAGAAAGCACTGTACTTACCTTACCGGGTCACTTT 235
 QY 61 ValThrSerHisLeuValAsnArgProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 236 CTGTCCAGAGGC-----AAGATGCCATGACGTCTTCATGATGATTCATTCATG 289
 QY 81 SerArgCysThrIleGlnGlnProIlePheAlaProAsnTyrIleLysGlyThrIleGln 100
 DB 290 AAAGACTGTGATGATCAAGACAGCCGTATTTGTGCAACTACATCAAGGACAGTGAAG 349
 QY 101 AlaAlaProGlyGlyGlyThrGluGlyGlnAlaValPheLysLeuSerPheArgLysGly 120
 DB 350 GCGAAGCGGAGAGTGGCTGGAGAGCTCTGCTTCCCTTCAAGTTACCTTACCGGAGG 409
 QY 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaAlaArgGly 139
 DB 410 GCGGCATTCATGATTCGACAGCGGATGCTCCAGGTGGATCTCAAGCTCCAGAGGTGA 469
 QY 140 IleProLeuGlySerValAsnTyrTrpPheArgThrSerGlyLeuTyrIleThrVal 159
 DB 470 GTCCCAAGTGAAGCCATAGTACTCTTACAGCCAGCGGSGCTATCTATGCTATCCCG 529
 QY 160 ProGlyAlaAlaValCysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
 DB 530 CCA-----GTGCGCAATGAAATGATACCTGCTCTGCGCTACCC----- 571
 QY 179 TyrGlyProProProGlyTyrThrValGlnProGlyGlyTyrGlyThrProProGlu 198
 DB 572 TATCCACCGCCCGCCACCTGAGTTC-----TATCCAGAGACCCCGCATG 613
 QY 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProPro-----MetGlyTyr 214
 DB 614 ATGAGCGGGGCGCATG-----GATATCGTGAAGCCCGCCACACCGCTTACCTGGGCGCC 667
 QY 215 GlyAlaProProVal---GlyTyrGlyValPro----- 224
 DB 668 ATGAACTTCGGTCAAGCGGCGCCGATGTCCTCTCCACTCTGAGCGGAGCAAGAGCC 727
 QY 225 ----- 225
 DB 728 GCAAGAGACCGCGCGCGCTTATCAACCCAGGCAATCTTCACAAGCTTACATGCCC 787
 QY 226 GlyGlyTyrGlyValProProGlyGlyTyrGly----- 236
 DB 788 ACGAGCGACCGCGCGCACTCTCACTACCAACCGAAGATGAAGACCAAGTAGGCC 847
 QY 237 -----ValProProGlyGly---TyrGlyAlaProPro----- 246
 DB 848 CTCGCGCTCCGCTCCACCTCCATCTCTTACCTTACCTCCATCGGAGGCTGG 907
 QY 247 GlyGlyTyrGlyValProProGlyGlyTyrGlyAlaProProGly-----Gly 262

RESULT 8
 ID AB267208 standard; cDNA, 1915 BP.
 XX
 AC AB267208;
 XX
 DT 26-MAR-2003 (first entry)
 XX
 DE Human secreted protein encoding cDNA SEQ ID NO 328.
 XX
 KW Human; secreted protein; nontropic; neuroprotective; cytostatic;
 KW virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV;
 KW viral; antibacterial; antiparkinsonian; antiaging; antianemic;
 KW antidiabetic; cancer; antineumatic; hepatotropic; cerebroprotective;
 KW antiinflammatory; antiallergic; antidiabetic; antitumor; anticonvulsant;
 KW antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine;
 KW cardiovascular disorder; neurological disease; nephrotropic;
 KW gene therapy; gene; de.
 XX
 OS Homo sapiens.
 XX
 PN NO200277186-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-US09188.
 XX
 PR 27-MAR-2001; 2001US-278650P.
 PR 12-SEP-2001; 2001US-0950082.
 PR 12-SEP-2001; 2001US-0950083.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM,
 XX
 DR WPI; 2003-040583/03.
 DR P-PSDB; ABP99787.
 XX
 PT New human secreted proteins encoded by genes contained in cDNA clones
 PT (e.g. HGAC13), useful for preventing, treating or diagnosing e.g.
 PT AIDS, multiple sclerosis, herpes virus, leukemia, tick-borne
 PT encephalitis or West Nile fever
 XX
 PS Claim 7; Page 1339; 2423pp; English.
 XX
 CC The invention relates to novel human genes (AB266891-AB268209) and the
 CC encoded secreted proteins (ABP9470-ABP9872) useful for preventing,
 CC treating or ameliorating medical conditions e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as
 CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,
 CC bacterial, fungal and parasitic infections.
 XX
 SQ Sequence 1915 BP; 401 A; 613 C; 481 G; 413 T; 7 other;
 Alignment Scores:
 Pred. No.: 5.15e-18 Length: 1915
 Score: 475.00 Matches: 132
 Percent Similarity: 47.08% Conservative: 37
 Best Local Similarity: 36.77% Mismatches: 113
 Query Match: 27.79% Indels: 77
 DB: 25 Gaps: 16
 US-09-864-291-5 (1-313) x AB267208 (1-1915)

Qy 1 MetAlaValaIaSerHisThrGluSerArgGlyAlaLeuIleProSerGlyGlu 20
 Db |||||.....
 59 ATGGCGCTCAACAGATCATCTCGAGGCGC---GGCGAGTATCTGTCAATACACCGGAG 115
 Qy 21 SerValLeuIleGlySerGluIleValaIaSerGlyPheLeuGlnIleProValGluSer 40
 Db |||||.....
 116 AGCATCTTAATGCTTATATATCATCGGAACTCATTAATGACATGAAAGACGTGCCA 175
 Qy 41 TyrLeuPheAsnGlyThrIleValaIaSerGlyPheLeuGlnIleProValGluSer 60
 Db |||||.....
 176 GAAGCTTCAAGAGGACCAAGAAAGGACATGCTTACCTTACCGGATCATCTT 235
 Qy 61 ValThrSerHisLeuValaIaSerProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db |||||.....
 236 CTGTCCAAAGGCGC---AAGATGCGCAAGCTCTTATATATGCAATTTATATCTCATG 289
 Qy 81 SerAspCysThrIleGlnIleProIlePheAlaProAsnThrIleGlyIleThrIleGln 100
 Db |||||.....
 290 AAGACTGTGAGTCAAGACCGCTTATTTGTGCAAACTCATCAAGGAAACATGAAAG 349
 Qy 101 AlaAlaProGlyGlyIleThrGluGlyGlnAlaValaIaPheLeuSerPheArgGly 120
 Db |||||.....
 350 GCGGAAGCGGAGATGCTGCGAAGCTCTGCTTCTTACAGTATCTTACCGGAGG 409
 Qy 121 GlyAlaIleGluPheAlaGlnLeuMetValaIaSerAlaIaSerAlaIaArgGly 139
 Db |||||.....
 410 GCGGCATTTGATTCGACAGAGCGATGCTGAGTGCATTCACCTTCAGAGGTGAA 469
 Qy 140 IleProLeuGlySerValaIaSerThrPheAsnThrSerGlyLeuThrIleIleThrVal 159
 Db |||||.....
 470 GTCCCAAGTGAAGCTTATGATCTTACATGACCAAGCGGAGCTTATGATATCCCGG 529
 Qy 160 ProGlyAlaAlaValaIaCysSerSerGlnThrProCys---ProAlaIleProIleValIle 178
 Db |||||.....
 530 CCA---GTCCCAATGAAGATGATACCTTCCCTGCGCTTACCGGCTTACCGGCTTACCGG 571
 Qy 179 TyrGlyProProProGlyIleThrValGlnProGlyGlyIleThrProProGly 198
 Db |||||.....
 572 TATCCACCGCGCCCACTGATGTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 613
 Qy 199 GlyTyrGlyAlaGlnProGlyIleThrValGlnProPro---MetGlyTyr 214
 Db |||||.....
 614 ATGACGAGGAGCATG---GATATGTCGACAGCCCAACACAGCGCTTACCTGAGGCC 667
 Qy 215 GlyAlaProProVal---GlyTyrGlyValPro--- 224
 Db |||||.....
 668 ATGGAACCTCCGCTCAGCGCGCCCGCATGTCCTTCACTCTGACAGCGGAGAGAGAGAGAG 727
 Qy 225 --- 225
 Db 728 GCAGAAAGACCGCGCAGCGCTTATCAACCCAGGCAATCTTCAACAGCTTACATGAGCC 787
 Qy 226 GlyIleTyrGlyValProProGlyIleThrVal--- 236
 Db 788 ACAGAGCAAGCGCGCGCATCTTCTTACCTTACCGGAGATGAAAGACCCAGTGAAGCC 847
 Qy 237 ---ValProProGlyIle---TyrGlyAlaProPro--- 246
 Db 848 CTCCTGCTCCCTGCTCCCTTCACTCTTACCTTACCGGAGAGAGAGAGAGAGAGAGAGAGAG 907
 Qy 247 GlyIleTyrGlyValProProGlyIleThrValAlaProProGly---Gly 262
 Db 908 CTGGGCGCTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 967
 Qy 263 TyrGlyAlaProProAlaGlyTyrGlyAlaProProAlaGlyAlaSerGlnAlaLeuPro 282
 Db 968 TACCAAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1024
 Qy 282 caaTyrGlnAlaProSerAlaGlyAsnThrAlaAlaSerHisArgSerMetThr 300
 Db 1025 CAGCTTCCATGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 1079

RESULT 9

AB870430
 ID AB870430 standard; cDNA, 1885 BP.
 XX
 AC AB870430;
 XX
 DT 27-NOV-2002 (first entry)
 XX
 DE Human bone remodelling gene #87.
 XX
 KM Bone remodelling; osteoporosis; human; gene; 88.
 XX
 OS Homo sapiens.
 XX
 PN US6426186-B1.
 XX
 PD 30-JUL-2002.
 XX
 PP 18-JAN-2000; 2000US-0484970.
 XX
 PR 18-JAN-2000; 2000US-0484970.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Jones KA, Volkmutz W, Walker MG,
 XX
 DR WPI: 2002-673014/72.
 XX
 XX A combination of polynucleotides which are co-expressed with genes
 PT known to be involved in bone remodeling and osteoporosis are useful in
 PT an array for the diagnosis of bone remodeling and osteoporosis
 PT associated disorders -
 XX
 PS Claim 1; Column 247-250; 206pp; English.
 XX
 CC The invention relates to a combination comprising a number of
 CC substantially purified and isolated polynucleotides which are
 CC co-expressed with genes known to be involved in bone remodeling and
 CC osteoporosis. The invention is used to diagnose disorders associated
 CC with bone remodeling or osteoporosis. AB870344-AB870512 represent
 CC human bone remodelling genes of the invention.
 XX
 SQ Sequence 1885 BP; 384 A; 610 C; 476 G; 415 T; 0 other;
 Alignment Scores:
 Pred. No.: 3.67e-17 Length: 1885
 Score: 459.50 Matches: 129
 Percent Similarity: 45.43% Conservative: 40
 Best Local Similarity: 34.68% Mismatches: 100
 Query Match: 26.89% Indels: 104
 DB: 24 Gaps: 17
 US-09-864-291-5 (1-313) x AB870430 (1-1885)
 QY 1 MetLAlaValaAngInSerHleThlGluSerHArgLylAlaLeuIleProSerGlyGlu 20
 DB 61 ATGGCGCTCAACAGAAATCACTCGAAGGC---GGCGAGTGAAGCTGCAATTAAGAGC 117
 QY 21 SerValLeuLysGlnGlyGluAspValAspLeuGlnLysProValGluSer 40
 DB 118 AGCATCTTAATGTCCTTATGATCAGTGAAGCTCACTTCAATTAAGAGAGAGAGAGC 177
 QY 41 TyrLeuPheAngInGlyThylLysGlyThylLeuPheLeuThylSerTyrArgValAlaPhe 60
 DB 178 GAAGCCTTCAAG 237
 QY 61 ValThrSerHleLeuValaAngInProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 238 CTGTCCAAAGGC-----AAGAGTCAATGATGATGATGATGATGATGATGATGATG 291
 QY 81 SerAspCysThrIleGlnInProIlePheAlaProAsnTyrIleLysGlyThylGln 100
 DB 292 AAAGACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 351

QY 101 AlaAlaProGlyLylGlyLylGlnGlnAlaValaPheLysLeuSerPheArgLylGly 120
 DB 352 GCGGAAGCGGAGAGTGGCTGGAGAGAGCTGCTCTTCTTCAAGATTAAGCTTCAAGGAG 411
 QY 121 GlyAlaIleGlnPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaAlaArgGly 139
 DB 412 GCGCCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 471
 QY 140 IleProLeuGlySerValaAngInTyrPheAspThrSerGlyLeuTyrIleLeuVal 159
 DB 472 GTCCCAAGTGAAGCTATGAGTCTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 531
 QY 160 ProGlyAlaAlaValaCysSerSerGlnThrProCys---ProAlaTyrProIleValIle 178
 DB 532 CCA-----GTCCCAATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 573
 QY 179 TyrGlyProProProProProProProProProProProProProProProProPro 185
 DB 574 TATCAACCGCCCGCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 633
 QY 186 TyrThrValGlnProGlyLylGlyLylGlnProProGlyLylGlyLylGlnProGly 205
 DB 634 TAC---GTGCAAGCC-----CCACACCGCCCTTAC-----CTGGG 666
 QY 206 GlyTyrGlyAlaPro----- 210
 DB 667 CCATGGAACCTCGGTTGACGCGCCCGCATGTCCTTCCCTCACTGACCGAGAGCA 726
 QY 210 ----- 210
 DB 727 GCGCGAG 786
 QY 211 -----PrometGlyTyrGlyAlaProProVal 219
 DB 787 GCCCAAG 846
 QY 220 GlyTyrGlyAlaProProGlyLylGlyLylGlnProProGlyLylGlyLylGlnProPro 239
 DB 847 GGC-----CCTCTTGGC-----TCCCTGCTTCCACCTTCACTTCACTTCACTTCA 891
 QY 240 GlyLylGlyLylAlaProProProGlyLylGlyLylGlnProProGlyLylGlyLylAlaPro 259
 DB 892 CCTTCCATGCGGAGCTGT 939
 QY 260 -----ProGlyLylGlyLylAlaProProAla 268
 DB 940 TGTTCCTCCCTCAAGTGTGATCATTAACAATTACCA-----GGAATACATTTGTG 990
 QY 269 GlyTyrGlyAlaProProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 288
 DB 991 GACCATTA---GGGCGCCCGGCTTGGAGAG---GTGCGCGCCAGCTTCCATCCAGAGCC 1046
 QY 289 AlaGlyAsnThrAlaAlaSerHleArgSerMetThr 300
 DB 1047 GAGAGCCCAAGT 1082
 RESULT 10
 ID AB84798
 AC AB84798;
 DT 14-AUG-2002 (first entry)
 XX
 DE Human cDNA differentially expressed in granulocytic cells #1369.
 XX
 KM Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KM viral infection; parasitic infection; protozoal infection;
 KM fungal infection; sterile inflammation; disease; psoriasis;
 KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KM cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KM adult respiratory distress syndrome; inflammatory bowel disease;
 KM Crohn's disease; ulcerative colitis; periodontal disease;

KM granulocyte activation; chronic inflammation; allergy.
 OS Homo sapiens.
 XX
 PN WO200228999-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 03-OCT-2001; 2001WO-US30821.
 XX
 PR 03-OCT-2000; 2000US-237189P.
 XX
 PA (GENE-) GENE LOGIC INC.
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX WPI; 2002-435328/46.
 DR
 PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 PS
 XX Claim 1; SEQ ID NO 1369; 114pp; English.
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) Gs by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating Gs; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease, also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published/pct_sequences.
 XX
 XX Sequence 220895 BP; 53760 A; 54808 C; 55831 G; 56496 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 8 72e-15 Length: 220895
 Score: 451.00 Matches: 84
 Percent Similarity: 63.64% Conservative: 7
 Best Local Similarity: 58.74% Mismatches: 10
 Query Match: 26.39% Indels: 42
 DB: 24 Gaps: 2

QY 174 TyrProIleValIleTyrGlyProProProProGlyTyrThrValGlnProGlyGlyTyr 193
 DB 16999 TTCACGATTAATGTCATATGAGAGCCCACTGACAGG-----TAT 170037
 QY 194 GLYThrProProGlyGlyTyrGlyValGlnProGlyGlyTyrGlyValProProMetGly 213
 DB 170038 GAGAGCCCACTTCCCGAATACGAGAGCCCACTGACGATATGAGAGCCCACTGATGAG 170097
 QY 214 TyrGlyValProProValGlyTyrGlyValProProGlyGlyTyrGlyValProProGly 233
 DB 170098 AATGAGAGCCCGCTGATGATACAGAGCTTACCTGCGATATGAGAGCCCACTGAT 170157
 QY 234 GLYTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProPro 253
 DB 170158 GATACGAGAGCCCACTGACGATATGAGAGCCCACTGATGATGAGAGCCCACT 170217
 QY 254 GLYGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValPro 273
 DB 170218 CTGAGATATGAGAGCCCACTGATATGAGAGCCCACTGATATGAGAGCCCA 170277
 QY 274 ProAGlyValAsnGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyVal 293
 DB 170278 CTGAGATATGAGAGCCCACTGATATGAGAGCCCACTGATATGAGAGCCCA 170337
 QY 294 AlaSerHisArgSerMetThrAlaGln-----GlnGlyThrSerLeuProThr 310
 DB 170338 AGGCTCAGAGATCTACAGAGCCCGCTCTGAAAAGAGGCTTCTCCCTCTGCG 170397
 QY 311 SerSerSer 313
 DB 170398 TCCTCTTCT 170406
 RESULT 11
 ID ABA67850 standard; DNA; 436 BP.
 XX ABA67850,
 AC
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human foetal liver single exon nucleic acid probe #16155.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00669.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000US-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-483447/52.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human foetal liver -
 XX
 PS Claim 4; SEQ ID NO 16155; 639pp + sequence listing; English.
 XX
 CC The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_jct_sequences.

XX Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Alignment Scores:

Pred. No.:	1.49e-16	Length:	436
Score:	438.00	Matches:	83
Percent Similarity:	62.86%	Conservative:	5
Best Local Similarity:	59.29%	Mismatches:	42
Query Match:	25.63%	Indels:	10
DB:	22	Gaps:	2

US-09-864-291-5 (1-313) x ABA67850 (1-436)

```

QY 177 ValIleTyGlyProProProProGlyTyThrValGlnProGlyGluTyGlyThrPro 196
   :::::|||||
DB 3 ATTGCTATGAGGCCCACTGCAGGA-----TATGAGCCCA 41

QY 197 ProGluGlyTyGlyValAlaGlnProGlyGlyTyGlyValAlaProPomeGlyTyGlyAla 216
   |||||
DB 42 CTTCCCGATGAGAGGCCCACTGCAGGATGAGGCCCACTGAGAAATGAGAGC 101

QY 217 ProProValGlyTyGlyValAlaProProGlyGlyTyGlyValAlaProProGlyGlyTyGly 236
   |||||
DB 102 CGCGCTGTGGATGAGAGGCCCACTGCAGGATGAGAGGCCCACTTGTGATACGA 161

QY 237 ValProProGlyGlyTyGlyValAlaProProGlyGlyTyGlyValAlaProProGlyGlyTy 256
   |||||
DB 162 GCCCACTGCAGGATGAGAGGCCCACTGCAGGATGAGAGGCCCACTTGTGATAT 221

QY 257 GlyAlaProProGlyGlyTyGlyValAlaProProAlaGlyTyGlyValAlaProProAlaGly 276
   |||||
DB 222 GGAACCCCACTGCGATGAGAGGCCCACTGCGATGAGAGGCCCACTGAGGA 281

QY 277 AengIuaIaLeuProProAlaTyGlyAlaProSerAlaGlyAserThAlaAserHis 296
   |||||
DB 282 AATGAGAGCCCGCTGCGATGAGAGGCCCACTGCGATGAGAGGCCCACTGAGGA 341

QY 297 ArgSerMetThrAlaGln-----GlnGluThrSerLeuProThrTherSerSer 313
   |||||
DB 342 GAATCTACAGCAGCCAGGCTCTGAAACGAGGCTTCTCTCCCTGCTGCTCTCTCT 401

```

RESULT 12

AAK42003
 ID AAK42003 standard; DNA; 436 BP.

AC AAK42003;

DT 06-NOV-2001 (first entry)

DB Human bone marrow expressed single exon probe SEQ ID NO: 16560.

KM Human; bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukemia; lymphoma; myeloma; ss.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0633366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI, 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -

XX Example 4; SEQ ID NO: 16560; 658pp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention.

XX Sequence 436 BP; 98 A; 146 C; 106 G; 86 T; 0 other;

Alignment Scores:

Pred. No.:	1.49e-16	Length:	436
Score:	438.00	Matches:	83
Percent Similarity:	62.86%	Conservative:	5
Best Local Similarity:	59.29%	Mismatches:	42
Query Match:	25.63%	Indels:	10
DB:	22	Gaps:	2

US-09-864-291-5 (1-313) x AAK42003 (1-436)

```

QY 177 ValIleTyGlyProProProProGlyTyThrValGlnProGlyGluTyGlyThrPro 196
   :::::|||||
DB 3 ATTGCTATGAGGCCCACTGCAGGA-----TATGAGCCCA 41

QY 197 ProGluGlyTyGlyValAlaGlnProGlyGlyTyGlyValAlaProPomeGlyTyGlyAla 216
   |||||
DB 42 CTTCCCGATGAGAGGCCCACTGCAGGATGAGGCCCACTGAGAAATGAGAGC 101

QY 217 ProProValGlyTyGlyValAlaProProGlyGlyTyGlyValAlaProProGlyGlyTy 236
   |||||
DB 102 CGCGCTGTGGATGAGAGGCCCACTGCAGGATGAGAGGCCCACTTGTGATACGA 161

QY 237 ValProProGlyGlyTyGlyValAlaProProGlyGlyTyGlyValAlaProProGlyGlyTy 256
   |||||
DB 162 GCCCACTGCAGGATGAGAGGCCCACTGCAGGATGAGAGGCCCACTTGTGATAT 221

QY 257 GlyAlaProProGlyGlyTyGlyValAlaProProAlaGlyTyGlyValAlaProProAlaGly 276
   |||||
DB 222 GGAACCCCACTGCGATGAGAGGCCCACTGCGATGAGAGGCCCACTGAGGA 281

QY 277 AengIuaIaLeuProProAlaTyGlyAlaProSerAlaGlyAserThAlaAserHis 296
   |||||
DB 282 AATGAGAGCCCGCTGCGATGAGAGGCCCACTGCGATGAGAGGCCCACTGAGGA 341

QY 297 ArgSerMetThrAlaGln-----GlnGluThrSerLeuProThrTherSerSer 313
   |||||
DB 342 GAATCTACAGCAGCCAGGCTCTGAAACGAGGCTTCTCTCCCTGCTGCTCTCTCT 401

```

RESULT 13

AAI48070
 ID AAI48070 standard; DNA; 436 BP.

AC AAI48070;

DT 17-OCT-2001 (first entry)

DB Probe #16756 used to measure gene expression in human placenta sample.

PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;

XX
 DR WPI; 2001-465570/50.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen
 PT is used in preventing, treating or ameliorating a medical condition -
 XX
 PS Disclosure; SEQ ID NO 7570; 1297pp + Sequence Listing; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a genomic sequence encoding a
 CC protein of the invention.
 XX
 SQ Sequence 7099 BP; 1782 A; 1545 C; 1445 G; 2327 T; 0 other;
 Alignment Scores:
 Pred. No.: 5,74e-15 Length: 7099
 Score: 429.50 Matches: 77
 Percent Similarity: 64.34% Conservative: 6
 Best Local Similarity: 59.69% Mismatches: 39
 Query Match: 25.13% Indels: 7
 DB: 22 Gaps: 1
 US-09-864-291-5 (1-313) x AML04882 (1-7099)
 QY 174 TyrProIleValIleTyrGlyProProProProGlyTyrThrValGlnProGlyGlyTyr 193
 DB 6719 TTCGAGTATTTGTTATGAGCCCACTGACAGA-----TAT 6757
 QY 194 GlyThrProProGlyGlyTyrGlyValAlaGlnProGlyGlyTyrGlyValAlaProProMetGly 213
 DB 6758 GAGGCCCACTCCCGATACGAGCCCACTGAGATATGAGCCCACTGAGAG 6817
 QY 214 TyrGlyAlaProProValGlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProProGly 233
 DB 6818 AATGAGGCCCGCTGTGATACAGACCTCACTGTGATATGAGCCCACTCTT 6877
 QY 234 GlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProPro 253
 DB 6878 GATACGAGGCCCACTGAGATATGAGCCCACTGAGATATGAGCCCACT 6937
 QY 254 GlyGlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProProAlaGlyTyrGlyValAlaPro 273
 DB 6938 CTGATATGAGAACCACTCTGATATGAGCCCACTCTGATATGAGCCCA 6997
 QY 274 ProAlaGlyValAlaProProAlaGlyTyrGlyValAlaProProAlaGlyTyrGlyValAlaPro 293
 DB 6998 CTGACAGAAATGAGGCCCGCTGCGGATACAGAGCTCACTGATGATCAGAGGCC 7057
 QY 294 AlaSerHisArgSerMetThrAlaGln 302
 DB 7058 AGGCTCAGAAATCTACAGACCCAG 7084

Search completed: December 16, 2003, 18:07:55
 Job time : 345.166 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: December 16, 2003, 17:36:35 / Search time 75.9653 Seconds
(Without alignments)
1818.631 Million cell updates/sec

Title: US-09-864-291-5
Perfect score: 1709
Sequence: 1 MAVNQSHTSRGALIPSGR.....ASHRSMTAQOETSLPTSSS 313

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DRV=xlp
-Q=/cgn2_1/USPTO.spool_p/US09864291/runat.15122003.160859.23332/app_query.fasta_1.846
-DB=Issued Patents NA -QPMT=fastap -SUPPLY=rm1 -MINMATCH=0.1 -LOOPTC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.ccl
-LIST=45 -DOCALLIGN=200 -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-MODE=LOCAL -OUTPMT=ptco -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09864291 @CEN 1.1 115 @runat.15122003.160859.23332 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOB=6 -DELEXT=7

Database: Issued Patents NA.*
1: /cgn2_6/prodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/prodata/2/ina/6C.COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	459.5	26.9	1885	4	US-09-484-9708-87 Sequence 87, Appl
2	227	13.3	2830	2	US-09-010-9288-1 Sequence 1, Appl
3	210.5	12.3	2824	3	US-09-010-9288-3 Sequence 3, Appl
4	210.5	12.3	4403765	2	US-09-103-840A-2 Sequence 2, Appl
5	210.5	12.3	4411529	3	US-09-103-840A-1 Sequence 1, Appl
6	206.5	12.1	654	3	US-08-998-416-1144 Sequence 1144, Ap
7	205.5	12.0	756	1	US-08-642-255-50 Sequence 50, Appl
8	198	11.6	2344	4	US-09-347-878-31 Sequence 31, Appl
9	198	11.6	2428	1	US-08-445-050-1 Sequence 1, Appl
10	198	11.6	2428	1	US-08-204-691-1 Sequence 1, Appl
11	198	11.6	2428	4	US-09-355-2958-2 Sequence 2, Appl
12	198	11.6	2487	3	US-08-370-223-12 Sequence 12, Appl

13	198	11.6	3018	1	US-08-347-718B-3 Sequence 3, Appl
14	198	11.6	3018	1	US-08-482-262-3 Sequence 3, Appl
15	198	11.6	3018	6	5200183-1 Patent No. 5200183
16	197.5	11.6	744	4	US-09-183-861-60 Sequence 60, Appl
17	197.5	11.6	744	4	US-09-022-765-60 Sequence 60, Appl
18	197.5	11.6	744	4	US-09-551-974A-60 Sequence 60, Appl
19	195	11.4	810	1	US-08-642-255-60 Sequence 60, Appl
20	193	11.3	432	1	US-08-642-255-48 Sequence 48, Appl
21	192.5	11.3	2824	2	US-09-010-928B-3 Sequence 3, Appl
22	189.5	11.1	11531	1	US-08-068-945B-1 Sequence 1, Appl
23	189.5	11.1	11531	1	US-08-442-806-1 Sequence 1, Appl
24	189.5	11.1	11531	4	US-09-355-2958-1 Sequence 1, Appl
25	178.5	10.4	1995	1	US-08-425-069-3 Sequence 3, Appl
26	178.5	10.4	1995	2	US-08-317-784B-3 Sequence 3, Appl
27	175.5	10.3	3747	2	US-09-080-897-1 Sequence 1, Appl
28	175.5	10.3	3747	3	US-09-323-735-1 Sequence 1, Appl
29	175	10.2	4403765	3	US-09-103-840A-2 Sequence 2, Appl
30	175	10.2	4411529	3	US-09-103-840A-1 Sequence 1, Appl
31	172	10.1	4378	2	US-09-080-897-3 Sequence 3, Appl
32	172	10.1	4378	3	US-09-323-735-3 Sequence 3, Appl
33	172	10.1	4399	3	US-08-899-595-2 Sequence 2, Appl
34	170	9.9	2210	3	US-08-464-700-1 Sequence 1, Appl
35	170	9.9	2242	1	US-08-641-627A-37 Sequence 37, Appl
36	170	9.9	2242	1	US-08-641-627A-37 Sequence 37, Appl
37	170	9.9	2427	2	US-08-678-039A-39 Sequence 39, Appl
38	168.5	9.9	5822	3	US-08-899-595-4 Sequence 4, Appl
39	168.5	9.9	5822	3	US-08-899-595-5 Sequence 5, Appl
40	167	9.8	3147	2	US-08-781-802-7 Sequence 7, Appl
41	167	9.8	3147	3	US-08-694-078-7 Sequence 7, Appl
42	167	9.8	3147	2	US-09-058-260-7 Sequence 7, Appl
43	167	9.8	35100	2	US-08-770-379-17 Sequence 17, Appl
44	167	9.8	35100	3	US-08-757-669A-17 Sequence 17, Appl
45	167	9.8	35100	4	US-09-230-371A-17 Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-484-9708-87
Sequence 87, Application US/09484970B
Patent No. 6426186
GENERAL INFORMATION:
APPLICANT: Jones, Karen A.
APPLICANT: Volkmuth, Wayne
APPLICANT: Walker, Michael G.
TITLE OF INVENTION: BONE REMODELING GENES
FILE REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 1885
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 232968.10CB1
US-09-484-9708-87

Alignment Scores:

Pred. No.: 1.05e-28 Length: 1885
Score: 459.50 Matches: 129
Percent Similarity: 45.43% Conservative: 40
Best Local Similarity: 34.68% Mismatches: 100
Query Match: 26.89% Indels: 104
DB: 4 Gaps: 17

US-09-864-291-5 (1-313) x US-09-484-9708-87 (1-1885)

Qy 1 MetAlaValAsnGlnSerHisrHrguSerArgGlyAlaLeuIleProSerGlygu 20
Db 61 ATGGCGCTCAACAGATCACTGGAGGCGC---GGCGAGTATCTCTCAATACCGAG 117


```

/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ TITLE OF INVENTION: TUBERCULOSIS
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, c, t, g or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 0.000385 Length: 4403765
Score: 210.50 Matches: 54
Percent Similarity: 56.00% Conservative: 27
Best Local Similarity: 54.00% Mismatches: 2
Query Match: 12.32% Indels: 17
Gaps: 10

US-09-864-291-5 (1-313) X US-09-103-840A-2 (1-4403765)

QY 183 ProProGlyYThrValGlnProGlyGlyThrProProGlyGlyYTrGlyAla 202
DB 4122678 CCGCTGGCTACCGCGTTGAACCA-----CCGCCGCGCGCGCGGATAT 4122722

QY 203 GlnProGlyGlyYTrGlyAlaProProMetGlyYTrGlyAlaProProValGlyYTrGly 222
DB 4122723 GCGCGCGCGGATAT-----CCGCC-----TACCCCGCTACACACCGCGGTACGAC 4122770

QY 223 ValProProGlyGlyYTrGlyValProProGlyGlyYTrGlyValProProGlyYTr 242
DB 4122771 -----CCGCCGCGGTATGATGCGCGCGCC--AGCTATGCG--CCCCCGCTGGCTAT 4122818

QY 243 GlyAlaProProGlyGlyYTrGlyValProProGlyGlyYTrGlyAlaProProGlyY 262
DB 4122819 GGT-----CCACCCCTCGGCTACCCCGCGCGCGCGCGGTGCGGCCACCGCC--GCG 4122872

QY 263 TyrGlyAlaProProAlaGlyYTrGlyAlaProProAlaGlyAlaGlnGlnAlaLeuProPro 282
DB 4122873 TATGCG-----CCACCGCTCGGCTATGCG--CCACCGGTGCGCGCGCGCGGTCAACCA 4122926

RESULT 5
US-09-103-840A-1
/ Sequence 1, Application US/09103840A
/ Patent No. 6294328
/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ TITLE OF INVENTION: TUBERCULOSIS
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 4411529
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```

Alignment Scores:
Pred. No.: 0.000386 Length: 4411529
Score: 210.50 Matches: 54
Percent Similarity: 56.00% Conservative: 27
Best Local Similarity: 54.00% Mismatches: 2
Query Match: 12.32% Indels: 17
Gaps: 10

US-09-864-291-5 (1-313) X US-09-103-840A-1 (1-4411529)

QY 183 ProProGlyYThrValGlnProGlyGlyThrProProGlyGlyYTrGlyAla 202
DB 4130483 CCGCTGGCTACCGCGTTGAACCA-----CCGCCGCGCGCGCGGATAT 4130527

QY 203 GlnProGlyGlyYTrGlyAlaProProMetGlyYTrGlyAlaProProValGlyYTrGly 222
DB 4130528 GCGCGCGCGGATAT-----CCGCC-----TACCCCGCTACACACCGCGGTACGAC 4130575

QY 223 ValProProGlyGlyYTrGlyValProProGlyGlyYTrGlyValProProGlyYTr 242
DB 4130576 -----CCGCCGCGGTATGATGCGCGCGCC--AGCTATGCG--CCCCCGCTGGCTAT 4130623

QY 243 GlyAlaProProGlyGlyYTrGlyValProProGlyGlyYTrGlyAlaProProGlyY 262
DB 4130624 GGT-----CCACCCCTCGGCTACCCCGCGCGCGCGCGGTGCGGCCACCGCC--GCG 4130677

QY 263 TyrGlyAlaProProAlaGlyYTrGlyAlaProProAlaGlyAlaGlnGlnAlaLeuProPro 282
DB 4130678 TATGCG-----CCACCGCTCGGCTATGCG--CCACCGGTGCGCGCGCGCGGTCAACCA 4130721

RESULT 6
US-08-998-416-1144
/ Sequence 1144, Application US/08998416
/ Patent No. 6239264
/ GENERAL INFORMATION:
/ APPLICANT: Philippaen, Peter
/ APPLICANT: Pohlmann, Rainer
/ APPLICANT: Steiner, Sabine
/ APPLICANT: Mohr, Christine
/ APPLICANT: Wendland, Jurgen
/ APPLICANT: Knechtel, Philipp
/ APPLICANT: Redischung, Corinne
/ TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
/ TITLE OF INVENTION: AND US8 THEROP
/ NUMBER OF SEQUENCES: 1152
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: No. 6239264artis Corporation
/ STREET: 3054 Cornwallis Road
/ CITY: Research Triangle Park
/ STATE: No. 6239264th Carolina
/ COUNTRY: USA
/ ZIP: 27709
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/998,416
/ FILING DATE: 24-DEC-1997
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: CH 0016/97
/ FILING DATE: 31-DEC-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Meigs, J. Timothy
/ REGISTRATION NUMBER: 38,241
/ REFERENCE/DOCKET NUMBER: PP/5-30306/A/CGC1976
/ TELEPHONE: 919-541-8587
/ TELEFAX: 919-541-8689
/ INFORMATION FOR SEQ ID NO: 1144:
/ SEQUENCE CHARACTERISTICS:
```


Db 308 CCCGAGACCCGGTGAACAGCCGCTCCAGGCGACCTGGCGGCGCTGGAGGTCCCGA 249
Qy 258 AAlaPro-----ProGlyGlyTyrGlyAlaPro-----ProAlaGlyTyrGlyAlaPro 273
Db 248 GACCCCGGTGAACAGCCGCTCCAGGCGACCTGGCGGCGCTGGAGGTCCCGA 189
Qy 274 -----ProAlaGlyAlaSerGlyAlaLeuProAlaTyrGlyAlaProSerAla---Gly 290
Db 188 GGTGACCAAGCCGCTCCAGGCGCA---CCTGGCGGCGCTGGAGGTCCCGA 132
Qy 291 AenThraAlaAlaSerHisArgSerMetThraAlaGlnGlnGlnThraSerLeuProThra 310
Db 131 GG-ACCAGCCGCTCCAGGCGCACTGGCGGCTCCAGGTCCCGAGCACCCGGTGA 73
Qy 311 SerSerSer 313
Db 72 ACCCGCTCC 64

RESULT 8

US-09-347-878-31
Sequence 31, Application US/09347878C
Patent No. 6376210
GENERAL INFORMATION:
APPLICANT: Yuan, Chong
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
FILE REFERENCE: 25885-1651
CURRENT APPLICATION NUMBER: US/09/347,878C
CURRENT FILING DATE: 1999-07-06
NUMBER OF SEQ ID NOS: 75
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 31
LENGTH: 2344
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (9)..(2252)
FEATURE:
OTHER INFORMATION: Human cholesterol esterase cDNA
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: M55201/Genbank
US-09-347-878-31

Alignment Scores:

Pred. No.:	3, 01e-07	Length:	2344
Score:	198.00	Matches:	110
Percent Similarity:	39.23%	Conservative:	32
Best Local Similarity:	30.39%	Mismatches:	127
Query Match:	11.59%	Indels:	94
DB:	4	Gaps:	26

US-09-864-291-5 (1-313) x US-09-347-878-31 (1-2344)

Qy 7 HisThraGlySerArgTyrGlyAlaLeuLeuProSerGlyGlySerValLeuGlnCys 26
Db 1176 TACACTGAGTCC-----TGGGCCGAGACCATCCAGAGAAATTAAGAAAGACTGTG 1229
Qy 27 GluAer-----ValaLeuLeuGlnCysPheLeuGlnLysProValGlu----- 39
Db 1230 GTGGACTTTGAGACCGATGTCTCTTCG---GTGCCACCGAGATGTCCTTACGCCAG 1286
Qy 40 -----SerTyrLeuPheAlaGlyThrLys 48
Db 1287 CACAGAGCCAAATGCCAAGTGCAGACCTACGCTTCTTTCCATCCCTCTCGG 1346
Qy 49 GlyThrLeuPheLeuThraSerTyrArgValValPheValThrSerHisLeuValaSer 68
Db 1347 ATGGCCGCTTAC-----CCCAATGGGTGGGGCGCAGCATGCATGAC 1391
Qy 69 PrometLeuSerPheMetMetProPheGlyLeuMetSer-----AaPys 83
Db 1392 ATTCAGTACGTTTGGGAAAGCCCTTGCACACCCGAGGCTACCGGCCCAAGACAG 1451

Qy 84 ThrIleGlnInProlIlePheAla-----ProAlaTyrIleGlyThrIleGlnAla 101
Db 1452 ACACTCTTAAGGCAATGATCGCTTACTGACACCAACTTTCGCAA-----ACAGG 1502
Qy 102 AAlaProGlyGly-----ThraGlyGlyAlaValaPheLys 114
Db 1503 GACCCCAATGGGCACTGGCTGGCCCAACACTGAGG-----CCC 1547
Qy 115 LeuSerPheArgGlyGlyGlyAlaIleGlnPheAlaLeuMetValaAlaSer 134
Db 1548 TACACTGAGAAACAGGCGTACTGAGATCAAGAAAGATG-----GGAGCAGC 1601
Qy 135 AAlaAlaAlaGlyIlePheLeuGlySerValaSerValaSerTyrPheArgThr--SerGly 153
Db 1602 TCCATGAAGCGGACCTGAGAACCAACTTCTGGCTACTGAGACCTTCACTTATCTGCG 1661
Qy 154 LeuTyrIleIleThraValProGlyAlaAla--ValCysSerSerGln----- 168
Db 1662 CTGCCCCAATGACCAAGACAGAGGCCACCTGTGTGCCCCCAAGAGGAGTCCGAGGCC 1721
Qy 169 -ThraProCysProAlaTyrProlIleValIleTyrGlyProProProGlyTyrThra 188
Db 1722 ACTCCCGTCC-----CCAGGCTGACTCCGAGACGCGCCCTGCGCCCAAGGT 1775
Qy 188 GlnProGlyGlyTyrGlyThrProProGlyGly--TyrGlyAlaGlnPro----- 204
Db 1776 GACTCCGGGGGCCCC-CCGTTGCCGCCCAAGGCTGACTCCGGGGCCCCCGTGGCGCC 1834
Qy 205 -----GlyGlyTyrGlyAlaProPromet-----GlyTyrGlyAlaProProva 219
Db 1835 CACGGGTGACTCCGGGGGCCCCCGGTGGCCCAAGGAGTCTCCGGGGCCCCCGGT 1894
Qy 219 GlnTyrGlyValProProGlyGlyTyrGlyValaProProGlyGlyTyrGlyValaProPr 239
Db 1895 G-----CGGCCAAGGTGACTCCGGGGCCCCCG-----GTGGCGCC 1933
Qy 239 G-----GlyGlyTyrGlyAlaProProGlyGlyTyrGlyValaProPro--GlyGlyTyrGly 257
Db 1934 CACGGGTGACTCCGGGGGCCCCCG-----GTGGCGGCCAAGGCTGACTCCGG 1981
Qy 257 ValaProPro-----GlyGlyTyrGlyAlaProProAlaGlyTyrGlyAlaPr 273
Db 1982 GCGCCCCCGTGGCCCGCCAGCGGTGACGCGGGGCCCCCG-----GTGCC 2029
Qy 273 GProAlaGlyAlaSerGlnAlaLeuProProAlaTyrGlyAlaProSerAlaGlyAlaThrAl 293
Db 2030 GCGCCAGGTGACTCCGGGGGCCCCCG-----GTGCCCGCCAGGCTGACTCCGG 2080
Qy 293 AAlaSerHisArgSerMetThraAlaGlnGlnGlnThraSer--LeuProThraThrSerSe 312
Db 2081 GCGCCCCCGTGAACCCCAAGGCTGACTCCGAGACGCGCCCGTGGCGCCCAAGGCTGA 2140
Qy 312 rSer 313
Db 2141 CTCC 2144

RESULT 9
US-08-445-050-1
Sequence 1, Application US/08445050
Patent No. 5763739
GENERAL INFORMATION:
APPLICANT: Blasekberg, Lars
APPLICANT: Edlund, Michael
APPLICANT: Hansson, Lennart
APPLICANT: Herrell, Olle
APPLICANT: Lundberg, Lennart
APPLICANT: Stromqvist, Mats
APPLICANT: Toernell, Jan
TITLE OF INVENTION: No. 5763739el Polypeptides
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: White & Case

STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: United States
 ZIP: 10036-2787
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/445,050
 FILING DATE:
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/204,691
 FILING DATE:
 APPLICATION NUMBER: SE 9300686-4
 FILING DATE: 01-MAR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: SE 9300722-7
 FILING DATE: 04-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Sterner Ph.D., Richard J
 REGISTRATION NUMBER: 35,372
 REFERENCE/DOCKET NUMBER: 1103326-850
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 819-8783
 TELEFAX: (212) 354-8113
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2428 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 TISSUE TYPE: mammary gland
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 82..2319
 OTHER INFORMATION: /product= "bile-salt-stimulated
 OTHER INFORMATION: lipase"
 FEATURE:
 NAME/KEY: exon
 LOCATION: 985..1173
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1174..1377
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1378..1575
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1576..2415
 FEATURE:
 NAME/KEY: mat_peptide
 LOCATION: 151..2316
 FEATURE:
 NAME/KEY: poly_a_signal
 LOCATION: 2397..2402
 FEATURE:
 NAME/KEY: repeat_region
 LOCATION: 1756..2283
 FEATURE:
 NAME/KEY: 5'UTR
 LOCATION: 1..81
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1756..1788

FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1789..1821
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1822..1854
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1855..1887
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1888..1920
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1921..1953
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1954..1986
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1987..2019
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2020..2052
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2053..2085
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2086..2118
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2119..2151
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2152..2184
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2185..2217
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2218..2250
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2251..2283
 US-08-445-050-1
 Alignment Scores:
 Pred. No.: 3,14e-07 Length: 2428
 Score: 198.00 Matches: 110
 Percent Similarity: 39.23% Conservative: 32
 Best Local Similarity: 30.39% Mismatches: 127
 Query Match: 11.59% Indels: 94
 DB: 1 Gaps: 26
 US-09-864-291-5 (1-313) x US-08-445-050-1 (1-2428)
 Qy 7 HistHrGluSerArgArgGlyAlaLeuIleProSerGlyGluSerValLeuLysGlnCys 26
 Db 1249 TACACCAAGTCC-----TGGCCCAAGACCCATCCAGAGAAATTAAGAAAGACTGTG 1302
 Qy 27 GluAsp-----ValAspLeuCysPheLeuGlnLysProValGlu----- 39
 Db 1303 GTGGACTTGTGAGACCATGCTCTTCTGTG---GTGCCCAAGAGATTGCCCTAGCCCAAG 1359
 Qy 40 -----SerTyrLeuPheAsnGlyThrLysLys 48
 Db 1360 CACAGAGCAATGCCAAGAGTGCAGACCTACGCTTCTGTTTCCCTATCCCTCTCGG 1419
 Qy 49 GlyThrLeuPheLeuThrSerTyrArgValValPheValThrSerHisLeuValAsnAsp 68
 Db 1420 ATGCCCGCTTAC-----CCCAATGGGTGGGGCCCAAGCAGTGCAGATGAC 1464

```

QY      69  ProMetLeuSerPheMetMetProPheGlyLeuMetSer-----ApcYs 83
Db      1465  ATTCAGTACGTTTGGGAGAGCCCTTCCACACCCCGCTACCGGCGCCACAGAG 1524
QY      84  ThrIleGlnProIlePheAla-----ProaenTYrIleIleGlyThrIleGlnAla 101
Db      1525  ACGAGTCTAAGGCCATGATGCTCTACGAGACCACTTGGCCAAA-----ACAGGG 1575
QY      102  AlaProGlyGlyGly-----TPrGlyGlyGlnAlaValPheYs 114
Db      1576  GACCCCAACATGGGCGACTCGGCTGCGCCACACACTGGAA-----CCC 1620
QY      115  LeuSerPheArgLyGlyGlyAlaIleGlnPheAlaGlnLeuMetValIleAlaIleSer 134
Db      1621  TACACTACGAAAGAACAGCGGCTACCTGAGATCACACAGAGATGAT-----GGCAGAGC 1674
QY      135  AlaAlaAlaArgGlyIleProIleuGlySerValaenTYrTPrPheAspThr---serGly 153
Db      1675  TTCATGAAAGCGGAGCTGAGAACTCACTTCCGCTGCTACTGAGACCTCTCATCTGGCG 1734
QY      154  LeuTYrIleIleThrValProGlyAlaAla-ValCysSerSerGln-----168
Db      1735  CTGCCCAACATGACCGACAGAGAGGCCACCCCTGTGCGCCCAACAGGGGACTCCAGAGCC 1794
QY      169  -ThrProCysProAlaTYrProIleValIleTYrGlyProProProGlyTYrTYrVa 188
Db      1795  ACTCCCGTGGCC-----CCACGGGTGACTCCGAGACCGCCCGTGGCGCCACGGGT 1848
QY      188  IGlInProGlyGlyuTYrGlyThnProProGluuIly---TYrGlyAlaGlnPro----- 204
Db      1849  GACTCGGGGGCCCCC-CCCGTCCCGCCACGGGTGACTCCGGAGCCCCCGCGTGGCGCC 1907
QY      205  ---GlyGlyTYrGlyAlaProProMet-----GlyTYrGlyAlaProProVa 219
Db      1908  CACGGGTGACTCCGGGGCCCCCCCCCGTGGCGCCACGGGTGACTCCGGGGCCCCCGCT 1967
QY      219  IGlTYrGlyValProProGlyGlyTYrGlyValProProGlyGlyTYrGlyValProPr 239
Db      1968  G-----CGGCCACGGGTGACTCCGGGGCCCCCCCC-----GTGGCGCC 2006
QY      239  o---GlyGlyTYrGlyAlaProProGlyGlyTYrGlyValProPro---GlyGlyTYrGl 257
Db      2007  CACGGGTGACTCCGGGGCCCCCCCC-----GTGGCGCCACGGGTGACTCCGG 2054
QY      257  ValaProPro-----GlyGlyTYrGlyAlaProProAlaGlyTYrGlyAlaPr 273
Db      2055  CGCCCCCGCGTGGCGCCACCGGGTACCGCGGGCCCCCCCC-----GTGCC 2102
QY      273  oProAlaGlyAenGlyAlaLeuProProAlaTYrGlyAlaProSerAlaGlyAenThrAl 293
Db      2103  GCCCAAGGTGACTCCGGGGCCCCCCCC-----GTGCCGCCACAGGGGTGACTCCGG 2153
QY      293  aAlaSerHlaArgSerMetThrAlaGlnGlnuThrSer---LeuProThrThrSerSe 312
Db      2154  GGCCCCCGCGTGGCGCCACCGGGTACCTCCGAGACCGCCCCCGCGTGGCGGTA 2213
QY      312  rSer 313
Db      2214  CTCC 2217

```

```

RESULT 10
US-08-204-691-1
Sequence 1, Application US/08204691
Patent No. 5827683
GENERAL INFORMATION:
APPLICANT: Blaeckberg, Lars
APPLICANT: Edlund, Michael
APPLICANT: Hansson, Lennart
APPLICANT: Herneil, Olie
APPLICANT: Lundberg, Lennart
APPLICANT: Stromqvist, Mats
APPLICANT: Toernell, Jan
TITLE OF INVENTION: No. 5827683el Polypeptides

```

```

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: White & Case
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/204,691
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9300686-4
FILING DATE: 01-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 9300722-7
FILING DATE: 04-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stenier Ph.D., Richard J
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-850
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)819-8783
FAX: (212)354-8113
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2428 base pairs
TYPE: nucleic acid
STRANDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: mammary gland
FEATURES:
NAME/KEY: CDS
LOCATION: 82..2319
OTHER INFORMATION: /product= "bile-salt-stimulated
OTHER INFORMATION: lipase"
FEATURES:
NAME/KEY: exon
LOCATION: 985..1173
FEATURES:
NAME/KEY: exon
LOCATION: 1174..1377
FEATURES:
NAME/KEY: exon
LOCATION: 1378..1575
FEATURES:
NAME/KEY: exon
LOCATION: 1576..2415
FEATURES:
NAME/KEY: mat_peptide
LOCATION: 151..2316
FEATURES:
NAME/KEY: poly_a_signal
LOCATION: 2397..2402
FEATURES:
NAME/KEY: repeat_region
LOCATION: 1756..2283
FEATURES:
NAME/KEY: 5'UTR
LOCATION: 1..81
FEATURES:
NAME/KEY: repeat_unit

```

```

: LOCATION: 1756..1788
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 1789..1821
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 1822..1854
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 1855..1887
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 1888..1920
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 1921..1953
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 1954..1986
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 1987..2019
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 2020..2052
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 2053..2085
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 2086..2118
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 2119..2151
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 2152..2184
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 2185..2217
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 2218..2250
: FEATURE:
: NAME/KEY: repeat unit
: LOCATION: 2251..2283
:
US-08-204-691-1

Alignment Scores:
Pred. No: 3.14e-07
Score: 198.00
Percent Similarity: 39.23%
Best Local Similarity: 30.39%
Query Match: 1 1.59%
Gaps: 26

US-09-864-291-5 (1-313) x US-08-204-691-1 (1-2428)
QY 7 HisThrGluSerArgArgGlyAlaLeuIleProSerGlyGluSerValIleuValGlnCys 26
Db 1249 TACGCCAGATCC-----TGGGCCGAGAGCCATCCGAGAAATTAAGAGAGACTGTG 1302
QY 27 GluAsp-----ValAspLeuCybPheLeuGlnLysProValGlu----- 39
Db 1303 GTGACCTTGTAGAGACGATGCTCTCTCTG---GTGCCACCGAGATTGCCCTAGCCGAG 1359
QY 40 -----SerrTyrlleuPheAsnGlyThrLysLys 48
Db 1360 CACAGAGCCATGCGAAGTGCAGACCTACCGCTTACCTGTTTCCATCCCTCTCGG 1419
QY 49 GlyThrIleuPheLeuThrSerrTyArgValAlaPheValThrSerrHileuValAsnAsp 68
Db 1420 ATGCCCGCTTAC-----CCCAATGGGTGGGGGCGGACCATGCATGCATGATGAC 1464

```

```

QY 69 ProMetIleuSerPheMetMetProPheGlyLeuMetSer-----AspCys 83
Db 1465 ATTCAGTACCTTTTCGGAAAGCCCTTGGCCACCCCGACGAGCTACCGGCCCAAGACAG 1524
QY 84 ThrIleGluGlnProIlePheAla-----ProMetTyrlleYsglyThrIleGlnAla 101
Db 1525 ACACTCTCTAAGGCCATGATCGCTTACTGACCACTTTCGCAAA-----ACAGGG 1575
QY 102 AlaProGlyGlyGly-----TrypIleGlyGlnAlaValPheLys 114
Db 1576 GACCCCAACATGGGGGACTCGGGCTGTGCCACACACTGGGAA-----CCC 1620
QY 115 LeuSerPheArgLysGlyGlyAlaIleGluPheAlaGlnIleuMetValLysAlaIleSer 134
Db 1621 TACACTACGAAACACAGCGCTACTCGAGATCAACAAAGAAATG-----GCCAGCAGC 1674
QY 135 AlaAlaAlaIleProGlyIleProLeuGlySerValAsnTyrlThrPheAspThr-----SerrGly 153
Db 1675 TCCATGAAAGCGAGACCTGAGAACCACTTCTGCGCTACTGAACTTCACTTATCTGGCG 1734
QY 154 LeuTyrlleIleThrValProGlyAlaAla-ValCybSerSerrGln----- 168
Db 1735 CTGCCCACAGTACCGACAGAGAGGCCACCTGTGTGCCCCCAAGGGAGACTCCAGGCC 1794
QY 169 -ThrProCybProAlaTyrlProIleValIleTyrglyProProProGlyTyrlThrVa 188
Db 1795 ACTCCCGTGGCC-----CCAGGGGTGACTCCGAGACCGCCCGTGGCCGACGAGGT 1848
QY 188 IGlNProGlyGlyTyrglyTyrlThrProProGlyGly-----TyrglyAlaGlnPro----- 204
Db 1849 GACTCCGGGGGCCCCC-CCCGTGCAGCCACAGGGTGACTCCGGGGGCCCCCGTGGCGGC 1907
QY 205 -GlyGlyTyrglyValaProProMet-----GlyTyrglyAlaProProVa 219
Db 1908 CAGGGTGACTCCGGGGGCCCCCGTGGCGGCCCAAGGGTGACTCCGGGGGCCCCCGGT 1967
QY 219 IGlTyrglyValaProProGlyGlyTyrglyValaProProGlyGlyTyrglyValaProPr 239
Db 1968 G-----CCGCCACGGGTGACTCCGGGGGCCCCCGCC-----GTGCGGCC 2006
QY 239 o-----GlyGlyTyrglyValaProProGlyGlyTyrglyValaProPro-----GlyGlyTyrgly 257
Db 2007 CAGGGTGACTCCGGGGGCCCCCGTGGCGGCCCAAGGGTGACTCCGGGGGCCCCCGGT 2054
QY 257 ValaProPro-----GlyGlyTyrglyValaProProAlaGlyTyrglyAlaPr 273
Db 2055 CGCCCCCGCGTGCAGCCGACAGGGTGAAGCGGGGCCCCCGCC-----GTGCC 2102
QY 273 oProAlaGlyAsnGlnAlaLeuProProAlaTyrglyAlaProSerrAlaGlyAsnThrAl 293
Db 2103 GCCCAGCGGGTGACTCCGGCGCCCCCGCC-----GTGCCGCCACAGGGTGACTCGG 2153
QY 293 AlaIleSerHisAluSerMetThrAlaGlnGlnGlnThrSerr-----LeuProThrThrSerrSe 312
Db 2154 GGGCCCCCGGTGACCCCAAGGGTGACTCCGAGACCGGCCCGTGCAGGCCAGGGTGA 2213
QY 312 rSer 313
Db 2214 CTCC 2217

RESULT 11
US-09-355-295B-2
Sequence 2, Application US/09355295B
Patent No. 6525241
GENERAL INFORMATION:
APPLICANT: Daltymple, M.
APPLICANT: Lundberg, L.
APPLICANT: Stromqvist, M.
TITLE OF INVENTION: Expression Methods
FILE REFERENCE: 1754 SEQUENCE LISTING v2a.txt
CURRENT APPLICATION NUMBER: US/09/355,295B
CURRENT FILING DATE: 1999-07-18

```

PRIOR APPLICATION NUMBER: PCT/SB99/00648
 PRIOR FILING DATE: 1999-04-21
 NUMBER OF SEQ ID NOS: 5
 SEQ ID NO 2
 LENGTH: 2428
 TYPE: DNA
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: CDNA
 FEATURE:
 OTHER INFORMATION: Mammary gland source
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 82..2319
 OTHER INFORMATION: /product= "bile salt-stimulated
 OTHER INFORMATION: lipase"
 FEATURE:
 NAME/KEY: exon
 LOCATION: 985..1173
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1174..1377
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1378..1575
 FEATURE:
 NAME/KEY: exon
 LOCATION: 1576..2415
 FEATURE:
 NAME/KEY: mat_peptide
 LOCATION: 151..2316
 FEATURE:
 NAME/KEY: polyA_signal
 LOCATION: 2397..2402
 FEATURE:
 NAME/KEY: repeat_region
 LOCATION: 1756..2283
 FEATURE:
 NAME/KEY: 5' UTR
 LOCATION: 1..81
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1756..1768
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1789..1821
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1822..1854
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1855..1887
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1888..1920
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1921..1953
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1954..1986
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 1987..2019
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2020..2052
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2053..2085
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2086..2118

FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2119..2151
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2152..2184
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2185..2217
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2218..2250
 FEATURE:
 NAME/KEY: repeat_unit
 LOCATION: 2251..2283
 US-09-355-295B-2
 Alignment Scores:
 Pred. No.: 3,14e-07 Length: 2428
 Score: 198.00 Matches: 110
 Percent Similarity: 39.23% Conservative: 32
 Best Local Similarity: 30.39% Mismatches: 127
 Query Match: 11.59% Indels: 94
 DB: 4 Gaps: 26
 US-09-864-291-5 (1-313) x US-09-355-295B-2 (1-2428)
 QY 7 HlEhTgJlUSeRfAgRgLyAlAlEulEProSeRgLyJlUSeRfAlEulEgInCyS 26
 DB 1249 TACACCGAGTCC-----TGGCCCAAGACCCATCCCAAGAAAGAAAGAAAGCTGTG 1302
 QY 27 GlUAp-----ValApLeuCySPhetLeuGInLySProValGIn----- 39
 DB 1303 GTGAGCTTGAACCGATGCTCTTCTG--GTGCCACAGAGATTGCCCTAGCCAG 1359
 QY 40 -----SerTyLeuPheAsnGlyThrLyS 48
 DB 1360 CACAGAGCAATGCAAGAGTGCAGACCTAGGCTACTGTTTCCATCCCTCGG 1419
 QY 49 GlyThrLeuPheLeuThSerTyArgValValPheValThrSerHleuValAsnAp 68
 DB 1420 ATGCCCGTCTAC-----CCCAATGGGTGGGGCCGACCAAGCAATATAC 1464
 QY 69 ProMetLeuSerPheMetSerProPheGlyLeuMetSer-----ApCyS 83
 DB 1465 ATTCACTACGTTTGGGAAGCCTTCCGCAACCCCAAGGCTACCGGCCCAAGACAG 1524
 QY 84 ThrIleGInProIlePheAla-----ProAsnTyrlElyGlyThrIleGInAla 101
 DB 1525 ACAGTCTTAAGCCATGATGCTGACCACTTGGCCAA-----ACAAGG 1575
 QY 102 AlAProGlyGly-----TProGlyGInAlaValPheLyS 114
 DB 1576 GACCCCAATGGGGAGCTGCGCTGCCCACACACTGGAA-----CCC 1620
 QY 115 LeuSerPheArgLyGlyValAlIleGInPheAlaGInLeuMetValLyAlaIleSer 134
 DB 1621 TACACTACGAAACAGCGGCTACTGAGATCAACCAAGATG-----GGCAGACG 1674
 QY 135 AlAlAlAlArgLyIleProLeuGlySerValAsnTyrlTriPheAspThr-----SerGly 153
 DB 1675 TCCATGAAGCGAGAGCTGAGACCAACTTCGCGCTACTGACCTCATCTATCTGGG 1734
 QY 154 LeuTyrlIleThValAProGlyAlAla-ValCySerSerGIn----- 168
 DB 1735 CTGCCCAAGTACCGACCAAGAGGCACACCTGTGCCCCCAAGGGGACTCCAGAGCC 1794
 QY 169 -ThrProCySProAlaTyrlProIleValIleTyrlGlyProProProGlyTyrlThrVa 188
 DB 1795 ACTCCCGTCCCC-----CAACGGTACTCCAGACCGCCCGCTGCGCCAGCGGT 1848
 QY 188 IGlAProGlyGInTyrlThrProProGly-----TyrlAlaGInPro----- 204

Db 1849 GACTCCGGGGCCCCC-CCGTCGCGCCACGCGGTGACTCCGGGGCCCCCCTGTCGCGC 1907
Qy 205 ----GlyGlyTyrglyAlaProProMet-----GlyTyrglyAlaProProVa 219
Db 1908 CACGGGTGACTCCGGGGCCCCCCTGTCGCGCCACGCGGTGACTCCGGGGCCCCCCTG 1967
Qy 219 1G1TyrglyValProProGlyGlyTyrglyValProProGlyGlyTyrglyValProPr 239
Db 1968 G-----CCGCCACGCGGTGACTCCGGGGCCCCC-----GTGCGCGC 2006
Qy 239 o----GlyGlyTyrglyAlaProProGlyGlyTyrglyValProPro----GlyGlyTyrgly 257
Db 2007 CACGGGTGACTCCGGGGCCCCC-----GTGCGCGCCACGCGGTGACTCCGG 2054
Qy 257 yAlaProPro-----GlyGlyTyrglyAlaProProAlaGlyTyrglyAlaPr 273
Db 2055 CCGCCCCCGCGCGCCACGCGGTGACCGCGGGCCCCC-----GTGCGC 2102
Qy 273 oProAlaGlyAlaGlyAlaLeuProProAlaTyrglyAlaProSerAlaGlyAlaThrAl 293
Db 2103 GCCCAGCGGTGACTCCGGGGCCCCC-----GTGCGCGCCACGCGGTGACTCCGG 2153
Qy 293 aAlaSerAlaArgSerMetThrAlaGlnGlnGlnThrSer--LeuProThrThrSerSe 312
Db 2154 GCGCCCCCGCGCGACCCCAAGGTGACTCCGAGACCGCCCCCGCGCGCCACGCGGTGA 2213
Qy 312 rSer 313
Db 2214 CTCC 2217

RESULT 12

US-08-370-223-12
Sequence 12, Application US/08370223
Patent No. 6107026

GENERAL INFORMATION:

APPLICANT: Lange, III, Louis G.

APPLICANT: Kumar, B. Vajsa

TITLE OF INVENTION: Methods and Reagents for RFLP Analysis

TITLE OF INVENTION: of the Human Cholesterol Esterase Gene

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSER: Allegretti & Witcoff, Ltd.

STREET: 10 South Wacker, Suite 3000

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/370,223

FILING DATE:

CLASSIFICATION: 307

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/053,308

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: McDonnell, John J.

REGISTRATION NUMBER: 26,949

REFERENCE/DOCKET NUMBER: 91,441

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-715-1000

TELEFAX: 312-715-1234

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 2487 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 104..2341
US-08-370-223-12

Alignment Scores:
Pred. No.: 3,24e-07 Length: 2487
Score: 198.00 Matches: 110
Percent Similarity: 39.23% Conservative: 32
Best Local Similarity: 30.39% Mismatches: 127
Query Match: 11.59% Indels: 94
Gaps: 26

US-09-864-291-5 (1-313) x US-08-370-223-12 (1-2487)

Qy 7 HisThrGluSerArgGlyAlaLeuIleProSerGlyGluSerValIleuTygGlnCys 26
Db 1271 TACACCGAGTCC-----TGGGCCACAGACCCATCCAGAGATTAAGAGAGACTGTG 1324
Qy 27 GluAP-----ValaAPLeuCyAPheLeuGlnIleAProValGlu----- 39
Db 1325 GTGACTTGTAGACCGAGTGTCTCTCTG---GTGCCACCGAGATTGCCCTAGCCGAG 1381
Qy 40 -----SerTyIleuPheanGlyThrIleuLys 48
Db 1382 CACAGAGCAATGCCAAGTGCACAGACTTACCTGTTCTTCCATCCCTCTCG 1441
Qy 49 GlyThrIleuPheLeuThrSerTyArgValIlePheValIleThrSerIleuValIleuAP 68
Db 1442 ATGCCCGCTAC-----CCCAATGGGGGGGGGGCCACCATCGAGATGAC 1486
Qy 69 ProMetLeuSerPheMetMetProPheGlyLeuMetSer-----APCys 83
Db 1487 ATTCAGTACGTTTCCGAGAGCCCTTCCGCCACCCCAAGGCTACCGGCCCAAGACAG 1546
Qy 84 ThrIleGlnIleProIlePheAla-----ProAntTyIleTyGlyThrIleGlnAla 101
Db 1547 ACAGTCTTAAGGCAATGATGCTTACTGACCAATCTTGCMAA-----ACAAGG 1597
Qy 102 AlaProGlyGly-----ThrGluGlnAlaValPheLys 114
Db 1598 GACCCCAACATGGGAGCTGCGTGCACACATGGAA-----CCC 1642
Qy 115 LeuSerPheArgGlyGlyAlaIleGluPheAlaIleuMetValIleAlaSer 134
Db 1643 TACACTACGGAACACGCGCTACTGAGATACCAAGAAATG-----GGCAGCAGC 1696
Qy 135 AlaAlaAlaArgGlyIleProLeuGlySerValAntTyITripPheAPThr---SerGly 153
Db 1697 TCATAGAAAGCGAGCTGAGAACCACTTCTGCGTACTGAGACCTCATTACTTGGCG 1756
Qy 154 LeuTyIleIleThrValProGlyAlaAla-ValCysSerSerGln----- 168
Db 1757 CTGCCACAGATGACGACCAAGAGCCACCCCTGTGCCCCCAGAGAGAGACTCGAGGCC 1816
Qy 169 -ThrProCyAProAlaTyPProIleValIleTyGlyProProProGlyTyThrVa 188
Db 1817 ACTCCCGTCCCC-----CACGGGTGACTCCGAGACCGCCCCCGTCCGCGCCAGGT 1870
Qy 188 1GlnProGlyGlyIleTyThrProProGlyGly---TyrglyAlaGlnPro----- 204
Db 1871 GACTCCGGGGCCCCC-CCGTCGCGCCACGCGGTGACTCCGGGGCCCCCCTGTCGCGC 1929
Qy 205 ----GlyGlyTyrglyAlaProProMet-----GlyTyrglyAlaProProVa 219
Db 1930 CACGGGTGACTCCGGGGCCCCCCTGTCGCGCCACGCGGTGACTCCGGGGCCCCCCTG 1989
Qy 219 1G1TyrglyValProProGlyGlyTyrglyValProProGlyGlyTyrglyValProPr 239
Db 1990 G-----CCGCCACGCGGTGACTCCGGGGCCCCC-----GTGCGCGC 2028


```

Db      2691 GCGGACGGGTGACTCCAGGCCCCCC-----GTGCACCACGGGTACTC CGG 2741
Cy      293 AAlSeRhIarSgSerMeThrAlaInglngJnrser--LeuProThrTsrse 312
          |||             :||         |::|||           |::::|
Db      2742 GGCCCCCCCcGGAgAcCcCAcGGGTGAcrccGaGAccccccGTgcCCGCCACGGGTGA 2801
          |||             rser 313
Cy              |||
Db            2802 CTCC 2805

RESULT 14
US-08-482-262-3
; Sequence 3, Application US/08482262
Patent No. 5821226
GENERAL INFORMATION:
APPLICANT: Wang, Chu-San
APPLICANT: Tang, Jordan J.N
TITLE OF INVENTION: BAL C-TAIL DRUG DELIVERY MOLECULES
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
City: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,262
FILING DATE: June 7, 1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRP151
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO.: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3018 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..742
OTHER INFORMATION: /function = "Nucleotides 679 through
OTHER INFORMATION: 2904 encode the amino acid sequence for the Human MlIk
US-08-482-262-3

Alignment Scores:
Pred. No.:                Length:        3018
Score:                    Matches:         110
Percent Similarity:       Conservative:    32
Best Local Similarity:   Mismatches:     127
Query Match:             Indels:         94
DB:                      Gaps:           26

US-09-864-291-5 (1-313) x US-08-482-262-3 (1-3018)

Cy      7 HisTrhgUsErArGrArgLylAlaleuilePrOsErgLyGlUsErValleuYsgIncys 26
          :::|||||||||         |||  |||||||  |||::|  |||
```

Dh	1937	TACAGCCGAGTCC-----TGGGCGCCAGAGCCCAATCCAGAGGAATTAGAGAAAGACTCTG	1895
Qy	27	GIuApP-----ValAapLeuCySpheLeuGlnLysProValGlu-----	39
Dh	1891	GTGGACTTTTGAGACCCGATGTCCTCTCTG---GTGCCACCGAGATATGCGCTAGCCAG	1947
Qy	40	-----SerTyrLeuPheAmGlnTyrThrLys	48
Dh	1948	CACAGAGCCAAATGCCAAGAGTGCAGAGACTTACAGCTACCTGTTTTCCATCCCTCCG	2007
Qy	49	GlyThrLeuPheLeuThrSerTyrArgValAlaPheValThrSerHISLeuValAsnAsp	68
Dh	2008	ATGCGCCGTAC-----CCCAATGGGTGGGGCGCCAGCATGCAAGTAC	2052
Qy	69	ProMetLeuSerPheMetMetProPheGlyLeuMetSer-----AspCys	83
Dh	2053	ATTCAAGTACGTTTTGCGAAGCCCTTGCCACCCCAAGGACTACCGGCCCAAGACAG	2112
Qy	84	ThrIleGlnGlnProIlePheAla-----ProAsnTyrIleLysGlyThrIleGlnAla	101
Dh	2113	ACAGTCTCTAAGGCGCATGATCGCTACTGACCAACTTGTGCAAA-----ACAGGG	2163
Qy	102	AlaProGlyGlnGly-----TrpGlnGlnAlaValAlaPheLys	114
Dh	2164	GACCCCAACATGGGCGACTGGCTGTGCCCCACACACTGGAA-----CCC	2208
Qy	115	LeuSerPheArgLysGlyValAlaIleGlnPheAlaGlnLeuMetValLysAlaAsp	134
Dh	2209	TACACTACGAAAAACAGCGGCTACCTGAGATCACCAAGAAATG-----GGCAGACG	2262
Qy	135	AlaAlaAlaArgGlyIleProLeuGlySerValAsnTyrTrpPheAspThr--SerGly	153
Dh	2263	TCCATGAGGAGAGCCTGAGAACCACTTCTGTGCTACTGAGACCTCATTACTGTGGG	2322
Qy	154	LeuTyrIleIleThrValProGlyAlaAla-ValCysSerSerGln-----	168
Dh	2323	CTGCCCAAGTACACCAAGAGGCGCACTGTGCCCCCACAGGAGACTCCAGAGCC	2382
Qy	169	-ThrProCysProAlaTyrProIleValIleTyrGlyProProProGlyTyrThrVal	188
Dh	2383	ACTCCCGTGGCCC-----CCAGCGGTGACTCGAGACCGGCCCGTGGCCGCAAGGAT	2436
Qy	188	IglnProGlyGlnTyrGlyThrProProGlyGly--TyrGlyAlaGlnPro-----	204
Dh	2437	GACTCCGAGGCGCCC-CCCGTGGCGCCACGCGGTGACTCGGAGGCGCCCCCGTGGCGCC	2495
Qy	205	-----GlyGlyTyrGlyAlaProPromet-----GlyTyrGlyAlaProProVal	219
Dh	2496	CACGCGGTGACTCGGGGCGCCCCCGGCGGCCACGAGGTGACTCGAGGCGCCCCCGGT	2555
Qy	219	IglnTyrGlyValProProGlyGlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProPr	239
Dh	2556	G-----CGCGCCACGAGGTGACTCGGGGCGCCCC-----GTGGCGCC	2594
Qy	239	G-----GlyGlyTyrGlyAlaProProGlyGlyTyrGlyValProPro-----GlyGlyTyrGly	257
Dh	2595	CACGCGGTGACTCGGGGCGCCCC-----GTGGCGCCACGAGGTGACTCGG	2642
Qy	257	ValAlaProPro-----GlyGlyTyrGlyAlaProProAlaIleTyrGlyAlaPr	273
Dh	2643	CGCCCCCGCGTGGCGGCCACGAGGTGACGCGGGGCGCCCC-----GTGGC	2690
Qy	273	OProlAlaGlyAsnGlnAlaLeuProProAlaTyrGlyAlaProSerAlaIleAsnThrAl	293
Dh	2691	GCCCAAGGTGACTCGGGCGCCCC-----GTGGCGCCACGAGGTGACTCGG	2741
Qy	293	AlaAspSerHISArgSerMetThrAlaGlnGlnGlnThrSer--LeuProThrThrSerSer	312
Dh	2742	GGGCGCCCGCGTACCCCAAGGAGTACTCGAGACGCGCCCGTGGCGCCACGAGGTGA	2801
Qy	312	rSer	313
Dh	2802	CTCC	2805

RESULT 15
 5200183-1
 / Patent No. 5200183
 / APPLICANT: TANG, JORDAN J.N.; MANG, CHI-SUN
 / TITLE OF INVENTION: RECOMBINANT BILE SALT ACTIVATED LIPASES
 / NUMBER OF SEQUENCES: 22
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/537,426
 / FILING DATE: 12-JUN-1990
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 504,635
 / FILING DATE: 04-APR-1990
 / APPLICATION NUMBER: 122,410
 / FILING DATE: 19-NOV-1987
 / SEQ ID NO: 1:
 / LENGTH: 3018
 5200183-1

Alignment Scores:
 Pred. No.: 4,14e-07 Length: 3018
 Score: 198.00 Matches: 110
 Percent Similarity: 39.23% Conservative: 32
 Best Local Similarity: 30.39% Mismatches: 127
 Query Match: 11.59% Indels: 94
 DB: Gaps: 26

US-09-864-291-5 (1-313) x 5200183-1 (1-3018)

QY 7 HieThqluSerArgaAglyAlaLeuileProSerGlylueValleuylGlnCyS 26
 DB 1837 TACACCGAGTCC-----TGAGCCAGACCCCAATCCAGAGATTAAGAAAGACTGTG 1890
 QY 27 GluAsp-----ValaLeuCySPhetleuGlnlyProValGlu----- 39
 DB 1891 GTGACCTTGAAGACCAATGCTCTCTG---GTGCCCAACAGATTGCCCTAAGCCAG 1947
 QY 40 -----SerTyrluPheAenGlyThrylYsLys 48
 DB 1948 CACAGAGCAATGCAAGATGCAAGACCAAGCTAAGCTGTTTCCCATCCCTCGG 2007
 QY 49 GlyThrluPheleuthSerTyArgValValPheValThrSerHleuValAenAsp 68
 DB 2008 ATGCCCGCTTAC-----CCCAATGGGTGGGGCCGACCATGCAATGAC 2052
 QY 69 PrometleuSerPheMetMetProPheGlyleuMetSer-----AspCyS 83
 DB 2053 ATTCACTAGCTTTTCGGAAAGCCTTGCACACCCCAAGGCTACCGGCCCCCAAGACAG 2112
 QY 84 ThrileGlnInProilePheAla-----ProAntTyrlleYsglyThrlleGlnAla 101
 DB 2113 ACGATCTTAAGGCAATGATCGCTACTGACCAACTTNGCAA-----ACAGGG 2163
 QY 102 AlaProGlylGly-----TrpGlnGlyGlnAlaValPheLys 114
 DB 2164 GACCCCAACATGGCGCACTCGGCTGTGCCACACACTGGGAA-----CCC 2208
 QY 115 LeuSerPheArglysglyAlaileGluPheleuMetVallysaAlaIaSer 134
 DB 2209 TACACTAGGAAAAAGCGGCTACTGAGATCACCAAGAAAGATG-----GGCAGCAGC 2262
 QY 135 AlaAlaIaArglyIleProleuGlySerValAntTyrlPheAenThr-----SerGly 153
 DB 2263 TCCATGAAGGAGGAGCTGAGAACCACTCTCGGCTACTGAGACCTCACTATCTGCGG 2322
 QY 154 LeuTyrlleIleThrValProGlyAlaAla-ValCySerSerGln----- 168
 DB 2323 CTGCCCAAGTGAACCGACAGAGGCCACCCCTGTGCCCCCAAGAGGAACTCCGAGGCC 2382
 QY 169 -ThrProCyProIaIaTyrlProIleValIleTyrglyProProProGlyTyThrVa 188
 DB 2383 ACTCCGAGTCCC-----CCACGGGTGACTCCGAGACCGCCCGTCCGCCCAAGGGT 2436

QY 188 lGlnProGlylGlyTyThrProProGlylGly-----TyrglyAlaGlnPro----- 204
 DB 2437 GACTCCGGGGCCCCC-CCCGTCCCGCCCAAGGATGATCCCGGGGCCCCCGATGCCGCC 2495
 QY 205 -----GlyGlyTyrglyAlaProProMet-----GlyTyrglyAlaProProVa 219
 DB 2496 CACGGGTGATCCGGGGGCCCCCGGTGCGGCCCAAGGATGATCCGGGGGCCCCCGGT 2555
 QY 219 lGlyTyrglyValProProGlylGlyTyrglyValProProGlylGlyTyrglyValProPr 239
 DB 2556 G-----CGGCCAAGGATGATCCCGGGGCCCCCGCC-----GTGCCGCC 2594
 QY 239 G-----GlyGlyTyrglyAlaProProGlylGlyTyrglyValProPro-----GlyGlyTyrgl 257
 DB 2595 CACGGGTGATCCCGGGGCCCCCGCC-----GTGCCGCCAAGGATGATCCCGG 2642
 QY 257 ValaProPro-----GlyGlyTyrglyAlaProProAlaGlyTyrglyAlaPr 273
 DB 2643 CGCCCCCGCGGCGGCCCAAGGATGATCCCGGGGCCCCCGCC-----GTGCC 2690
 QY 273 OProAlaGlyAanGlnAlaLeuProProAlaTyrglyAlaProSerAlaGlyAanThrAl 293
 DB 2691 GCCCAAGGATGATCCCGGGGCCCCCGCC-----GTGCCGCCAAGGATGATCCCGG 2741
 QY 293 AlaSerHleuArgSerMetThrlaGlnGlnIuThrSer---leuProThrThrSerSe 312
 DB 2742 GGCCTCCCGCGTGAACCCCAAGGATGATCCGAGACCGCCCGCGTCCGCCCAAGGATGA 2801
 QY 312 rSer 313
 DB 2802 CTCC 2805

Search completed: December 16, 2003, 23:51:58
 Job time : 11:55 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame_p1n model

Run on: December 16, 2003, 17:58:55 / Search time 306.146 Seconds
(without alignments)
3398.008 Million cell updates/sec

Title: US-09-864-291-5
Perfect score: 1709
Sequence: 1 MAVNQHSHSRRCALIPSGR.....ASHRSMTAQOETSLPTSSS 313

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2201672 seqs, 166179599 residues
Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+g2n.model -DEV=x1p
-Q=/cg2_1/USPRO.epool.p/US09844291/runat.15122003.160900.23360/app.query.fasta.1.846
-DB=Published Applications NA -OPT=fastap -SUFFIX=emb -MINMATCH=0.1
-LOOPCT=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blom62
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODB=LOCL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09844291@cgn.1.1.108@runat.15122003.160900.23360
-NCPU=6 -ICPU=3 -NO MMAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSBLOCLCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database: Published Applications NA:
1: /cg2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cg2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cg2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cg2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cg2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cg2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cg2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cg2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cg2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cg2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cg2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cg2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cg2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cg2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cg2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cg2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cg2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cg2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES		
Result No.	Query	Description
Score	Match Length	ID

1	543	31.8	467	11	US-09-764-891-1230	Sequence 1230, Ap
2	438	25.6	436	9	US-09-864-761-31561	Sequence 31561, Ap
3	429.5	25.1	7099	11	US-09-764-891-7570	Sequence 7570, Ap
4	353.5	20.7	432	11	US-09-918-995-33536	Sequence 33536, A
5	347	20.3	321	11	US-09-764-891-1356	Sequence 1356, Ap
6	320.5	18.8	409	10	US-09-960-352-63984	Sequence 63984, Ap
7	241	14.1	593	13	US-10-027-633-290832	Sequence 290832, A
8	241	14.1	593	14	US-10-027-633-290832	Sequence 290832, A
9	239	14.0	511	13	US-10-029-386-24981	Sequence 24981, A
10	218	12.8	1743	10	US-09-887-576-788	Sequence 788, App
11	216	12.6	2016	10	US-09-938-842A-1004	Sequence 2004, Ap
12	215	12.5	1812	10	US-09-925-300-724	Sequence 724, App
13	214	12.5	2176	13	US-10-007-926A-215	Sequence 215, App
14	211	12.3	439	11	US-09-918-995-4554	Sequence 4554, Ap
15	209.5	12.3	2709	15	US-10-156-761-3543	Sequence 3543, Ap
16	209.5	12.3	9025608	15	US-10-156-761-1	Sequence 1, Appl
17	206	12.1	511	13	US-10-029-386-24981	Sequence 24981, A
18	206	12.1	1674	15	US-10-156-761-3052	Sequence 3052, App
19	206	12.1	9025608	15	US-10-156-761-1	Sequence 1, Appl
20	205.5	12.0	1909	13	US-10-094-748-1166	Sequence 1166, Ap
21	205.5	12.0	2709	15	US-10-156-761-3543	Sequence 3543, Ap
22	204	11.9	541	13	US-10-029-386-11265	Sequence 11265, A
23	203	11.9	3760	10	US-09-887-576-458	Sequence 458, App
24	202	11.8	534	10	US-09-938-842A-968	Sequence 968, App
25	202	11.8	731	9	US-09-770-149-103	Sequence 103, App
26	198	11.6	2428	10	US-09-969-347-220	Sequence 220, App
27	198	11.6	2428	11	US-09-418-176-1	Sequence 1, Appl
28	198	11.6	3004	9	US-09-873-367C-446	Sequence 446, App
29	197.5	11.6	744	13	US-09-874-923-60	Sequence 60, Appl
30	197.5	11.6	744	10	US-09-991-496-60	Sequence 2538, Ap
31	195.5	11.4	3138	15	US-10-156-761-2538	Sequence 3570, Ap
32	193	11.3	1869	15	US-10-156-761-3570	Sequence 1092, Ap
33	191	11.2	540	10	US-09-938-842A-1092	Sequence 647, App
34	190	11.1	80557	16	US-10-080-170-647	Sequence 2531, Ap
35	184	10.7	2430	15	US-10-156-761-2531	Sequence 7526, Ap
36	182.5	10.7	1733	15	US-10-128-714-7526	Sequence 7527, Ap
37	182	10.6	291	11	US-09-764-891-7521	Sequence 263, App
38	182	10.6	507	11	US-09-770-961-263	Sequence 2170, Ap
39	178.5	10.4	1958	10	US-09-880-107-2170	Sequence 207, App
40	178.5	10.4	1958	13	US-10-007-926A-207	Sequence 118, App
41	178.5	10.4	1958	13	US-10-210-120-118	Sequence 225, App
42	178.5	10.4	2188	10	US-09-962-832-225	Sequence 19944, A
43	178.5	10.4	2188	13	US-09-814-351-19944	Sequence 52, Appl
44	178.5	10.4	2468	13	US-10-210-120-52	Sequence 9949, Ap
45	178.5	10.4	8664	15	US-10-198-846-9949	

ALIGNMENTS

RESULT 1
US-09-764-891-1230
Sequence 1230, Application US/09764891
Publication No. US2003007808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1230
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (421)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1230
Alignment Scores:

Pred. No.: 2.28e-46 Length: 467
 Score: 543.00 Matches: 104
 Percent Similarity: 69.46% Conservative: 12
 Best Local Similarity: 62.28% Mismatches: 38
 Query Match: 31.77% Indels: 13
 DB: 11 Gaps: 4
 US-09-864-291-5 (1-313) x US-09-764-891-1230 (1-467)

QY 118 ArglySGlygYAlaIleGlnpHealagInleuMetVallyeAlaIleSerAlaIaIa 137
 DB 3 AGAAATGAGAGGCAATTAATTCACCACTGATGAGGAAAGCTGCTGCTGCTGCC 62
 QY 138 ArgGlyIleProleuGlySerValaIleTyrTrpPheagThrSerGlyLeuTyIleIle 157
 DB 63 CGAGATTCCTCCATTGAACTTAATGACTGCTTCACTCTGATGAGATTTATTAAT 122
 QY 158 ThrValProGlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleVal 177
 DB 123 ACTGGGGAAGGG---AATATGCACTCCACAGATGCTGT-----TCAGTTAT 170
 QY 178 IleTyGlyProProProProGlyTyThrValGlnProGlyIleTyGlyThrProPro 197
 DB 171 GTCATGAGAGCCCACTGACAGA-----TATGAGCCCACT 209
 QY 198 GluGlyTyGlyAlaGlnProGlyTyGlyAlaProProMetGlyTyGlyAlaPro 217
 DB 210 CCGGATAGGAGCCCACTGACAGATATGAGCCCACTGAGAAATGAGAGCCCG 269
 QY 218 ProValGlyTyGlyValaProProGlyTyGlyValaProProGlyTyGlyVala 237
 DB 270 CCTGGGATACAGAGCCCACTGATGAGATATGAGAGCCCACTTGAATACGAGAGCC 329
 QY 238 ProProGlyTyGlyValaProProGlyTyGlyValaProProGlyTyGlyVala 257
 DB 330 CCACCTGAGATATGAGAGCCCACTGAGATATGAGAGCCCACTTGTGATATGAG 389
 QY 257 ValaProProGlyTyGlyValaProProAlaGlyTyGlyAlaProProAlaGly 277
 DB 390 AACCCCACTCTGAGATATGAGAGCCCACTTGCATATGAGAGCCCACTGACGAGAA 449
 QY 277 ngIuaIaleuProProAla 283
 DB 450 TGAAGGCC---CCGCTTGC 465

RESULT 2
 US-09-864-761-31561
 ; Sequence 31561, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; FILE REFERENCE: Aecmics-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263,6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: US 60/236,667
 ; PRIOR FILING DATE: 2000-09-21
 ; PRIOR APPLICATION NUMBER: US 09/608,408
 ; PRIOR FILING DATE: 2000-06-30
 ; PRIOR APPLICATION NUMBER: US 09/774,203
 ; PRIOR FILING DATE: 2001-01-29
 ; NUMBER OF SEQ ID NOS: 49117
 ; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 31561
 ; LENGTH: 436
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO 299716.4
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 9.7
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.9
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
 ; OTHER INFORMATION: EST HUMAN HIT: AM169980.1, EVALU8 8.00e-11
 ; OTHER INFORMATION: SWISSPROT HIT: P18616, EVALU8 3.00e-13
 ; US-09-864-761-31561

Alignment Scores:
 Pred. No.: 1.02e-35 Length: 436
 Score: 438.00 Matches: 83
 Percent Similarity: 62.86% Conservative: 5
 Best Local Similarity: 59.29% Mismatches: 42
 Query Match: 25.63% Indels: 10
 DB: 9 Gaps: 2

US-09-864-291-5 (1-313) x US-09-864-761-31561 (1-436)

QY 177 ValIleTyGlyProProProProGlyTyThrValGlnProGlyIleTyGlyThrPro 196
 DB 3 ATTGCTATATGAGAGCCCACTGACAGA-----TATGAGAGCCCA 41
 QY 197 ProGluGlyTyGlyAlaGlnProGlyTyGlyValaProProMetGlyTyGlyAla 216
 DB 42 CTTCCGATACGAGAGCCCACTGACAGATATGAGAGCCCACTGAGAAATGAGAGCC 101
 QY 217 ProProValGlyTyGlyValaProProGlyTyGlyValaProProGlyTyGlyVala 236
 DB 102 CCGCTGGGATACAGAGCCCACTGATGAGATATGAGAGCCCACTTGAATACGAG 161
 QY 237 ValProProGlyTyGlyValaProProGlyTyGlyValaProProGlyTyGlyVala 256
 DB 162 GCCCACTGAGATATGAGAGCCCACTTGAATATGAGAGCCCACTTGAATAT 221
 QY 257 GluAlaProProGlyTyGlyValaProProAlaGlyTyGlyAlaProProAlaGly 276
 DB 222 GAAACCCCACTCTGAGATATGAGAGCCCACTTGCATATGAGAGCCCACTGACAGA 281
 QY 277 AangIuaIaleuProProAlaTyrGluAlaProSerAlaGlyAanThrAlaAlaSerHis 296
 DB 282 AATGAGAGCCCGCTGCGGATACAGAGCCCACTGATGAGAGAGCCGCTCAG 341
 QY 297 ArgSerMetThrAlaGln-----GlnGluThrSerIleuProThrThrSerSer 313

```

Db      342 GAATCTACAGCAGCCGACCCTCGTAAAGAAGCCTTCCTTGCTGGCTCTTTCT 401
          ||| |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| 
RESULT 3
US-09-764-891-7570
; Sequence 7570, Application US/09764891
; Publication No. US2003007780BA1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764, 891
PRIORITY FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7570
LENGTH: 7099
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-7570

Alignment Scores:
Pred. No.: 2, 4e-33 Length: 7099
Score: 429.50 Matches: 77
Percent Similarity: 64.34% Conservative: 6
Best Local Similarity: 59.69% Mismatches: 39
Query Match: 25.13% Indels: 7
DB: 11 Gaps: 1

US-09-864-291-5 (1-313) x US-09-764-891-7570 (1-7099)

Qy      174 TyrProIleValIleTyrglyProProProProglyTYrThrValGlnProglyglutyr 193
       ::::::::::::::::::::-----TAT 675
Db      6719 TTCCAAGTAATGTCTAATGAAGGCCCACTGCAGAG-----
Qy      194 GlyHrProProGluGlyTYrglyAlaGlnProglyglYTYrglyAlaProPrometgyl 213
       GGAGGCCCAACTCCCGAGATACAGAACCCCACTGAGATATGAGAGCCCAACCGTAGGA 681
Db      6758
Qy      214 TyrglyAlaProProvalGlyTYrglyValProProglyglYTYrglyValProProgly 233
       6818 AATGAAGCCCGCGCTGTGGATACAGAGCTCACCTGTGCAGATATGAGAGCCCACTCTT 687
Db      6818
Qy      234 GlyTYrglyValProProglyglYTYrglyAlaProProglyglYTYrglyValProPro 253
       6878 GGATACGAGAGCCCACCTCGCAGAGATATGAGAGCCCACTCTAGATATGAGAGCCCACT 693
Db      6878
Qy      254 GlyglYTYrglyAlaProProglyglYTYrglyAlaProProalaglyTYrglyAlaPro 273
       6918 CTTCGATATGAAACCCCACTCTCGAGATATGAGAGCCCACTCTCGATATGAGAGCCCCA 699
Db      6918
Qy      274 ProIlaGlyAmGluAlaLeuProProalaTYrglyAlaProseAlaGlYAsnThrzla 293
       6998 CCTGAGGAAATGAAGGCCCGCGCTGCGGAGATACAGAGCTCACCTGCTGAGATCAGAGACC 705
Db      6998
Qy      294 AlaSerHisArgSerMetThrLagin 302
       7058 AGGCTCAGGAATCTACAGCAGGCCGAG 7084
Db      7058

RESULT 4
US-09-918-995-33536
; Sequence 33536, Application US/09918995
; Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918, 995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235, 076
PRIOR FILING DATE: 1999-01-20

```

```

NUMBER OF SEQ ID NOS: 38054
SOFTWARE: PageSeq for Windows Version 3.0
SEQ ID NO 33536
LENGTH: 432
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: mlec_feature
LOCATION: (1)...(432)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-33536

Alignment Scores:
Pred. No.: 4.01e-27 432
Score: 353.50 Matches: 71
Percent Similarity: 68.35% Conservative: 24
Best Local Similarity: 51.08% Mismatches: 41
Query Match: 20.68% Indels: 3
DB: 11 Gaps: 2

US-09-864-291-5 (1-313) x US-09-918-995-33536 (1-432)
QY 1 Metalavalanqlnserthsrthrgluserargayglalalaleuileproserdylglu 20
DB 18 ATGGCGCTCAACAGATCATCTCGAGGGC---GGCGAGTGAATGTCAATACACCGAG 74
QY 21 SerValleuylsglnCysglusapValaApleuCyshleuGlnlyProValGlsuer 40
DB 75 AGCATCTTAATGCGCTTAATATCATCGTGAACCTCAATCAATGAATGAAGACGTGCA 13
QY 41 TyrlauphelaqllyThrylysglyThryleuphelaThrseryrArgValAlphe 60
DB 135 GAAGCCTCAAGAGGACCAAGAAAGGCACTGCTCACTTACCCTTACCGGGTATCTTT 19
QY 61 ValThrserthsrleuValasapPrometleuSerPheMetMetProPheGlyleuMet 80
DB 195 CTGTCGAAGGGC-----AAGATGCCATGCAGTCTCTCATATCATATTTATCTCATG 24
QY 81 SerAspCyethrllleGlnProIlePheAlaProAsnTyrllelysglyThrllleGln 100
DB 249 AAGACCTGTAGATCAAGCAGCCCGATTTGTTGCAAACTACATCATAGGAAACGTGAAG 300
QY 101 AlAlaAProGlyGlyTyrlGlnGlnAlaValPheYalSeuSerPheArglysgly 120
DB 309 GCGGAAGCGGAGAGTGGCTGGGAAGGCTGCTCTTCAAGTTACCTTCAAGGCAAGG 360
QY 121 GYAlAlAlleGlnPheAlaGlnleuMetVallyAlaAlaIleSerAlAlAlaArgly 139
DB 369 GGGGCATTAAGTTGCGAAGCGGATGCTCCAGGTGAGCATCTCAAGCTCCAGAGGT 425

RESULT 5
US-09-764-891-1356
Sequence 1356, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
PRIOR FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1356
LENGTH: 321
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (321)

```

OTHER INFORMATION: n equals a,c,g, or c
US-09-764-891-1356

Alignment Scores:

Pred. No.:	1.27e-26	Length:	321
Score:	347.00	Matches:	69
Percent Similarity:	81.90%	Conservative:	17
Best Local Similarity:	65.71%	Mismatches:	18
Query Match:	20.30%	Indels:	1
DB:	11	Gaps:	0

US-09-864-291-5 (1-313) x US-09-764-891-1356 (1-321)

Qy 11 ArgArgGlyAlaLeuIleProSerGlyGluSerValLeuIleGlyGlnCysGluAlaProValAlaP 30
Db 3 CGCGCGGAGCCCTCAATCCCTTAACGGTGAAGCTCTCTGAAGCGCTCCGAAATGTGGAG 62
Qy 31 LeuCysePheLeuGlnIleProValGluSerTyrLeuPheAlaGlyTyrIleGlyGlyTyr 50
Db 63 CTCTCCCTCCCAAGGATCAGAAAGGCTCAATGTCTTGTAGGTAGAAAGACAGAGACA 122
Qy 51 LeuPheLeuTherTyrAlaGlyValPheValTherIleLeuValAlaAspProme 70
Db 123 TTGTTCTCACTTCATACCGGAGATTTCATTAATTCATGCTCCATCACTGATCCCATG 182
Qy 71 LeuSerPheMetSerProPheGlyLeuMetSerAspCysThrIleGluGlnProIlePhe 90
Db 183 TTGTTCTTATGATGCAATTCATTCATGACAACTCTGTTGAACACAGATATT 242
Qy 91 AlaProAsnTyrIleGlyGlyTyrIleGlnAlaAlaProGlyGlyTyrGluGly 110
Db 243 GCGCAAACTTCATTAAAGGACATTCAGGACGCTCAATATGGKGGCTGGAAAGACA 302
Qy 110 AlaValPheLeu 114
Db 303 AGCTACTTTTAA 315

RESULT 6

US-09-960-352-6984
Sequence 6984, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Ningling
APPLICANT: Byatt, John C.
APPLICANT: Mathiasagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 6984
LENGTH: 409
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 30-LIB3058-024-Q1-K1-H9
US-09-960-352-6984

Alignment Scores:

Pred. No.:	8.56e-24	Length:	409
Score:	320.50	Matches:	64
Percent Similarity:	67.18%	Conservative:	24
Best Local Similarity:	48.85%	Mismatches:	40
Query Match:	18.75%	Indels:	3
DB:	10	Gaps:	2

US-09-864-291-5 (1-313) x US-09-960-352-6984 (1-409)

Qy 1 MetAlaValaLeuGlnSerHleThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 23 ATGGCTCTCATCAGAACCACTCGAGAGGC--GGCGAGGTGATCTGCACCAACACCGAG 79

Qy 21 SerValLeuIleGlnCysGluAlaProValAlaLeuIleGlyGlnCysGluAlaProValAlaP 40
Db 80 AGCATCTCTGATGTCATGACCATGATGATGATGATGATGATGATGATGATGATGATGATG 139
Qy 41 TyrLeuPheAlaGlyTyrIleGlyGlyTyrIlePheLeuTherTyrAlaGlyValPhe 60
Db 140 GAGGCTTCAGAGGAGCAAGAAAGCAACCGTCTTACCTTACCCGTCAGGTCATCTT 199
Qy 61 ValThrSerIleLeuValAlaAspPromeLeuSerPheMetMetProPheGlyLeuMet 80
Db 200 CTGTCCAG-----GCGAGGATCCCATGCACTCTTCATGATGATGATGATGATGATG 253
Qy 81 SerAspCysThrIleGlnGlnProIlePheAlaProAsnTyrIleGlyGlyTyrIleGln 100
Db 254 AAGACTGTGATGATCAGACAGCTGTGTGTGGGCAAACTACATCATCAAGGACAGTGAAG 313
Qy 101 AlaAlaProGlyGlyTyrGluGlyGlnAlaValPheLeuLeuSerPheArgGly 120
Db 314 GCGAAGCAGAGAGGAGGCTGGGAGGCTGTGATCTCAAGTTGACCTTATGTTGCGG 373
Qy 121 GlyAlaIleGluPheAlaGlnLeuMetValLeu 131
Db 374 GCGGCAATGATGATGACAGCGGATGTACAG 406

RESULT 7

US-10-027-632-290832
Sequence 290832, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 290832
LENGTH: 593
TYPE: DNA
ORGANISM: Human
US-10-027-632-290832

Alignment Scores:

Pred. No.:	1.68e-15	Length:	593
Score:	241.00	Matches:	73
Percent Similarity:	47.40%	Conservative:	18
Best Local Similarity:	38.02%	Mismatches:	65
Query Match:	14.10%	Indels:	36
DB:	13	Gaps:	10

US-09-864-291-5 (1-313) x US-10-027-632-290832 (1-593)

Qy 45 GlyThrIleGlyGlyTyrLeuPheLeuTherTyrArgValAlaPheValTherSerHis 64
Db 48 GGGACCAAAAGGACCACTTACCTTACCCCTTC-----ATCTTCTGTCCAG---- 98
Qy 65 LeuValAlaAspPromeLeuSerPheMetMetProPheGlyLeuMetSerAspCysThr 84

Db 99 ----GCCAGAGATGCCAGAGCTCTCGTAATGTCATTATCTTGAAGGCTATGAG 155
 QY 85 11GAGUGlnProIlePheAlaProAntyTTrIleYgIlyThr11GlnAlaAlaProGly 104
 Db 156 ATCAACAGAGCTGTGTGTTGACAAACATGATCAAGGAAACAGTGAACCGAAGAGAGCC 215
 QY 105 G1YGLYTPGlnGlyGlnAlaValPheYLeuSerPheArgYgIlyGlyAla11Gln 124
 Db 216 GGTGCTGGAGAGCTCTGCTTCGCGACAGTCCGCTTCACGGCAGAGGGGCGCACTGAA 275
 QY 125 PheAlaGlnLeuMetValIlyAlaAlaSerAlaAlaAlaArgY11leProLeuGlySer 144
 Db 276 TCTGACAGAGAGTGTCTCAGAGTGAATCATCTCAAGCGTCAAGGGGTGAAGCCCTC 329
 QY 145 ValAntyTTrPheAspThrSerGlyLeuTTrIleIleThr---ValProGlyAlaAla 163
 Db 330 -----AGTGAAGCTACGCGCACTCTCAAGCTGCTGCTGCAAGACA 368
 QY 164 -----ValCySerSerGlnThrProCyS---ProAlaTyrProIle 176
 Db 369 TGTGCTTTCCCGCGCAGTCACTAGTGAATGTATCTGCTGCTGCTGCTGCTGCT 425
 QY 177 ValIleTyrGlyProProProGlyTyrThrValGlnProGlyGlyIlyTyrGlyThrPro 196
 Db 426 -----TCTCACACACCTTCCCGCTGAGTCTATCCAGGA 461
 QY 197 ProGlnGlyTyrGlyAlaGlnProGlyIlyTyrGlyAlaProPro-----Met 212
 Db 462 CCCAAGATGACAGAGGCCATG-----GGGTATGTGCAACCCCGCACAGCTGCTGTCT 515
 QY 213 G1YTYrGlyAlaProProValGlyTyrGlyValPro 224
 Db 516 GGGCCCATGGAACCTCCGCTGACGCCCAAGATGTTCC 551

RESULT 8

US-10-027-632-290832
 / Sequence 290832, Application US/10027632
 / GENERAL INFORMATION:
 / APPLICANT: Wang, David G.
 / TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
 / TITLE OF INVENTION: Polymorphisms in the Human Genome
 / FILE REFERENCE: 108827.129
 / CURRENT APPLICATION NUMBER: US/10/027,632
 / CURRENT FILING DATE: 2002-04-30
 / PRIOR APPLICATION NUMBER: US 60/218,006
 / PRIOR FILING DATE: 2000-07-12
 / PRIOR APPLICATION NUMBER: US 60/198,676
 / PRIOR FILING DATE: 2000-04-20
 / PRIOR APPLICATION NUMBER: US 60/193,483
 / PRIOR FILING DATE: 2000-03-29
 / PRIOR APPLICATION NUMBER: US 60/185,218
 / PRIOR FILING DATE: 2000-02-24
 / PRIOR APPLICATION NUMBER: US 60/167,363
 / PRIOR FILING DATE: 1999-11-23
 / PRIOR APPLICATION NUMBER: US 60/156,358
 / PRIOR FILING DATE: 1999-09-28
 / PRIOR APPLICATION NUMBER: US 60/146,002
 / PRIOR FILING DATE: 1999-08-09
 / NUMBER OF SEQ ID NOS: 325720
 / SOFTWARE: FASTSEQ for Windows Version 4.0
 / SEQ ID NO 290832
 / LENGTH: 593
 / TYPE: DNA
 / ORGANISM: Human
 / US-10-027-632-290832

Alignment Scores:

Pred. No.: 1,68e-15 Length: 593
 Score: 241.00 Matches: 73
 Percent Similarity: 47.40% Conservative: 18
 Best Local Similarity: 38.02% Mismatches: 65
 Query Match: 14.10% Indels: 36
 Db: 14 Gaps: 10

US-09-864-291-5 (1-313) x US-10-027-632-290832 (1-593)

QY 45 G1YThrYLeYgIlyThrLeuPheLeuThrSerTyrArgValAlaPheValThrSerHis 64
 Db 48 GGGACCAAGAAAGGACCACTTACCTTACCCCTTC-----ATCTTCTGTCAGG--- 98
 QY 65 LeuValAntyAspPrometLeuSerPheMetMetProPheGlyLeuMetSerAspCysThr 84
 Db 99 ----GCCAAGATGCCACGCACTCTCGTAATGTCATTATCTTGAAGGCTATGAG 155
 QY 85 11GAGUGlnProIlePheAlaProAntyTTrIleYgIlyThr11GlnAlaAlaProGly 104
 Db 156 ATCAACAGAGCTGTGTGTTGACAAACATGATCAAGGAAACAGTGAACCGAAGAGAGCC 215
 QY 105 G1YGLYTPGlnGlyGlnAlaValPheYLeuSerPheArgYgIlyGlyAla11Gln 124
 Db 216 GGTGCTGGAGAGCTCTGCTTCGCGACAGTCCGCTTCACGGCAGAGGGGCGCACTGAA 275
 QY 125 PheAlaGlnLeuMetValIlyAlaAlaSerAlaAlaAlaArgY11leProLeuGlySer 144
 Db 276 TCTGACAGAGAGTGTCTCAGAGTGAATCATCTCAAGCGTCAAGGGGTGAAGCCCTC 329
 QY 145 ValAntyTTrPheAspThrSerGlyLeuTTrIleIleThr---ValProGlyAlaAla 163
 Db 330 -----AGTGAAGCTACGCGCACTCTCAAGCTGCTGCTGCAAGACA 368
 QY 164 -----ValCySerSerGlnThrProCyS---ProAlaTyrProIle 176
 Db 369 TGTGCTTTCCCGCGCAGTCACTAGTGAATGTATCTGCTGCTGCTGCTGCTGCT 425
 QY 177 ValIleTyrGlyProProProGlyTyrThrValGlnProGlyGlyIlyTyrGlyThrPro 196
 Db 426 -----TCTCACACACCTTCCCGCTGAGTCTATCCAGGA 461
 QY 197 ProGlnGlyTyrGlyAlaGlnProGlyIlyTyrGlyAlaProPro-----Met 212
 Db 462 CCCAAGATGACAGAGGCCATG-----GGGTATGTGCAACCCCGCACAGCTGCTGTCT 515
 QY 213 G1YTYrGlyAlaProProValGlyTyrGlyValPro 224
 Db 516 GGGCCCATGGAACCTCCGCTGACGCCCAAGATGTTCC 551

RESULT 9

US-10-029-386-24981
 / Sequence 24981, Application US/10029386
 / Publication No. US20030194704A1
 / GENERAL INFORMATION:
 / APPLICANT: Penn, Sharon G.
 / APPLICANT: Rank, David R.
 / APPLICANT: Hanzel, David K.
 / TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 / TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
 / FILE REFERENCE: ABOWICA-X-2
 / CURRENT APPLICATION NUMBER: US/10/029,386
 / CURRENT FILING DATE: 2001-12-20
 / NUMBER OF SEQ ID NOS: 34288
 / SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 / SEQ ID NO 24981
 / LENGTH: 511
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 / FEATURES:
 / OTHER INFORMATION: MAP TO 297205.1
 / OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.48
 / US-10-029-386-24981

Alignment Scores:

Pred. No.: 2.23e-15 Length: 511
 Score: 239.00 Matches: 51
 Percent Similarity: 45.97% Conservative: 6
 Best Local Similarity: 41.13% Mismatches: 59
 Query Match: 13.98% Indels: 8


```

1  APPLICANT: Kieps, Joel
2  APPLICANT: Wang, Xun
3  APPLICANT: Zhu, Tong
4  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
5  TITLE OF INVENTION: SAME, AND METHODS OF USE
6  FILE REFERENCE: SCRIPI300-3
7  CURRENT APPLICATION NUMBER: US/09/938,842A
8  CURRENT FILING DATE: 2001-08-24
9  PRIOR APPLICATION NUMBER: US 60/227,866
10 PRIOR FILING DATE: 2000-08-24
11 PRIOR APPLICATION NUMBER: US 60/264,647
12 PRIOR FILING DATE: 2001-01-16
13 PRIOR APPLICATION NUMBER: US 60/300,111
14 PRIOR FILING DATE: 2001-06-22
15 NUMBER OF SEQ ID NOS: 5379
16 SEQ ID NO 2004
17 LENGTH: 2016
18 TYPE: DNA
19 ORGANISM: Arabidopsis thaliana
20 US-09-938-842A-2004

```

Alignment Scores:			
Pred. No.:	2.69e-12	Length:	2016
Score:	216.00	Matches:	42
Percent Similarity:	54.44%	Conservative:	7
Best local Similarity:	46.67%	Mismatches:	37
Query Match:	12.64%	Indels:	4
DB:	10	Gaps:	2

Oy	210	ProPheMetCylYrGlyAlaProProGlnIgLtyrGlyValProProGlyIgLtyrGly	229
		: : : :	
Db	466	CcACCAAGTAGTCAT-----CCACCCTTAAACAAG-----CCTCAGGGGGCGCTAAGA	513
Oy	230	ValProProGlyIgLtyrGlyValProProGlyIgLtyrGlyValaProProGlyIgLtyr	249
Db	514	GTAACCTCCAAGCTGGCATAGGAAGATCAAGACTGGGTATTGAGATCCTCAAGCGGGTTAT	573
Oy	250	GlyValProProGlyIgLtyrGlyAlaProProGlyIgLtyrGlyValaProProAlaGly	269
		: : : :	
Db	574	GGAAATACCTCAAGCTGGCATAGGAAGTACTCAAGCGGGTATGAAATCCTCAAGTGAGC	633
Oy	270	TyrGlyAlaProProAlaGlyAsnGluAlaLeuProProAlaTyrglyAlaProSerAla	289
		: : : :	
Db	634	TATGGAATGCCCAAGTGGTCTTCCAGAAAGCTTGATGAAGCAATGCACAATGAGTTGAG	693
Oy	290	GlyAsnThrAlaAlaSerHisArgSerMet	299
		: : : :	
Db	694	GGCTTAAGCTTGAAGATTTGAATCAAG	723

```

RESULT 12
US-09-925-300-724
: Sequence 724, Application US/09925300
: Patent No. US20020151681A1
: GENERAL INFORMATION:
: APPLICANT: Craig Rosen,
: APPLICANT: Steve Ruben,
: TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA101
: CURRENT APPLICATION NUMBER: US/09/925,300
: CURRENT FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05988
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1890
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 724
: LENGTH: 1812
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-925-300-724

```

Alignment Scores:	
Pred. No.:	2,98e-12
Score:	215.00
Percent Similarity:	44.83%
Best Local Similarity:	39.08%
Query Match:	12.58%
DB:	10
Length:	1812
Matches:	68
Conservative:	10
Mismatches:	64
Indels:	32
Gaps:	12

```

US-09-864-291-5 (1-313) x US-09-925-300-724 (1-1812).
Qy 160 ProGlyValAlaValAlaCysSerSerGlnThrProCys-----177
   |||||::::::
Db 3 CCCGGCTCCATCTTGGGAGAGACCGGGTGGGCTGGAGCTGGCTGGGGGTGAGATG 62
   |||||:::
Qy 172 -----ProAlaIleProIleValIleTyrGlyProProProProGlyTyrThrValGln 188
   |||||:::
Db 63 TCATACCCAGGCGATATCCCCCAAGGCTAC---CCACCTTTCCTGGATATCTCTCGCA 115
   |||||:::
Qy 190 ProGlyValTyrGlyTyrThrProProGlnGlyTyrGlyValAlaGlnProGlyValTyrGlyVala 208
   |||||:::
Db 120 GGTCAAGAGATCATCTTTTCCCTTCCTGTGTAGATATCTTATCTCAAGGGCTTT-----177
   |||||:::
Qy 210 ProProMetGlyTyrGlyVala-----ProProValGlyTyrGlyValaProPro 223
   |||||:::
Db 174 CTTCCATATGGAGAGAGGTGCTTACCCACAGACAGTGGCCAGAGATGGGTAC-----CCA 22
   |||||:::
Qy 226 GlyValTyrGlyValaPro---ProGlyValTyrGlyValaProProGlyValTyrGlyVala 244
   |||||:::
Db 225 GAGAGCTGAGGCTTACCTTCGGCTGGAGGTATATCCAGGCCCTGGAGGCTATCTGGTGGC 28
   |||||:::
Qy 245 Pro---ProGlyValTyr-----GlyValaProProGly---GlyTyrGlyVala 258
   |||||:::
Db 285 CCACAGCCCAAGGGGAGACTCATCTTATCCCGAGATTTCTTCCAGGCGCAGAGATTTGAGATC 344
   |||||:::
Qy 259 ProProGlyValTyrGlyValaProProAlaGlyTyrGlyValaProProAlaGlyValaGln 276
   |||||:::
Db 345 CCACCAAGTGGAGCAAGGCTTT-----TCTGGGTATCCACAGCCACTTCAACAAGCTTAT 396
   |||||:::
Qy 279 AlaLeuProProAlaTyrGlnAlaProSerAlaGlyValaMetThrAlaSerHisArgSer 299
   |||||:::
Db 399 GGAGGTGGTCCAGCA---CAGGTTTCCATACCTGGATGGCTTTCCTGGAGAGA-----CAG 444
   |||||:::
Qy 299 MetThrAlaGlnGlnGlnIleThrSerLeuProThrThrSerSer 312
   |||||:::
Db 450 ATGCTTCTCAGATATCTGGAGAGCAACTTACTTACCTTAACT 491

```

```

/ RESULT 13
/ US-10-007-926A-215
/ Sequence 215, Application US/10007926A
/ Publication No. US20030143539A1
/ GENERAL INFORMATION:
/ APPLICANT: BERTUCCI, FRANCOIS
/ APPLICANT: HOULGATTE, REMI
/ APPLICANT: BIRNBAUM, DANIEL
/ APPLICANT: NGUYEN, CATHERINE
/ APPLICANT: VIENS, PATRICE
/ APPLICANT: FERT, VINCENT
/ TITLE OF INVENTION: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS
/ TITLE OF INVENTION: USING ARRAYS OF CANDIDATE GENES
/ FILER REFERENCE: 1546-R-00
/ CURRENT APPLICATION NUMBER: US/10/007,926A
/ PRIOR FILING DATE: 2001-12-07
/ PRIOR APPLICATION NUMBER: 60/254,090
/ NUMBER OF SEQ ID NOS: 468
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 215
/ LENGTH: 2176
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: annexin a7 (ANXA7) gene.

```

US-10-007-926A-215

Alignment Scores:	
Pred. No.:	4,73e-12
Score:	214.00
Percent Similarity:	47.3%
Best Local Similarity:	42.76%
Query Match:	12,52%
DB:	13
Gaps:	11
Length:	2176
Matches:	65
Conservative:	7
Mismatches:	58
Indels:	22
Gaps:	11

US-09-864-291-5 (1-313) X US-10-007-926A-215 (1-2176)

[illegible]

RESULT 14

```

US-09-918-995-4554
? Sequence 4554, Application US/09918995
? Publication No. US2003007362A1
? GENERAL INFORMATION:
? APPLICANT: Hyaseq, Inc.
? TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
? TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
? FILE REFERENCE: 20411-756
? CURRENT APPLICATION NUMBER: US/09/918,995
? CURRENT FILING DATE: 2001-07-30
? PRIOR APPLICATION NUMBER: US/09/235,076
? PRIOR FILING DATE: 1999-01-20
? NUMBER OF SEQ ID NOS: 38054
? SOFTWARE: FastSeq for Windows Version 3.0
? SEQ ID NO 4554
? LENGTH: 439
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-918-995-4554

```

Alignment Scores:	
Pred. No.:	1.3e-12
Score:	211.07
Percent Similarity:	50.77%
Best Local Similarity:	46.55%
Query Match:	12.35%
DB:	11
Length:	433
Matches:	60
Conservative:	6
Mismatches:	44
Indels:	20
Gaps:	10

US-09-864-291-5 (1-313) X US-09-918-995-4554 (1-439)

```

Dh      49  CCAAGGTAATCCCCCAACAAAGGCTAC---CCACTTTCCCTGGATATCTCTCGCAGGTGAG 105
Qy      172  ProAlaATyProIleValIleTyrGlyProProProGlyTyrTyrValAlaGlnProGly 151
Dh      106  GAGTCATCTTTTCCCCCTTCCTGGTCAGATACCTTATCTTATCTAGTGGCTTT-----CTCCCA 159
Qy      192  GluTyrGlyTyrProProGlyGlyTyrGlyValaGlnProGlyGlyTyrGlyValaProPro 211
Dh      106  GAGTCATCTTTTCCCCCTTCCTGGTCAGATACCTTATCTTATCTAGTGGCTTT-----CTCCCA 159
Qy      212  MetGlyTyrGlyVala-----ProProValGlyTyrGlyValaProProGlyGly 227
Dh      160  ATGGAGAGAGAGGCTCACCACCAAGTCGCAAGTAGTGCTAC-----CCAGAGCT 210
Qy      228  TyrGlyValaPro---ProGlyGlyTyrGlyValaProProGlyGlyTyrGlyValaPro--- 245
Dh      211  GGAAGCTACCTCCGCGCTCGAGGTTATCTCAACCCCTGAGGCTATCCTGATGCCCAAG 270
Qy      246  ProGlyGlyTyr-----GlyValaProProGly---GlyTyrGlyValaProPro 266
Dh      271  CCAGGGGGAGCTCATCTCATCTCCGAGGTTCTCTCAAGGCCAAGGATTTGGAGTCCCA 330
Qy      261  GlyGlyTyrGlyValaProProAlaGlyTyrGlyValaProProAlaGlyValaGlnGlyVala 280
Dh      331  GGTGGAGCAGGGCTTT-----TCGGGTATCCACAGCCACTTCACAGTCTTATGAGGT 384
Qy      281  ProProAlaTyrGlyValaProSerAlaGly 290
Dh      385  GGTCCAGCA---CAGTTCACACTCTGAT 411

```

RESULT 15
RIS-10-156

[illegible]

Alignment Scores:	
Pred. No.:	1,78e-11
Score:	209.50
Percent Similarity:	36.24%
Best Local Similarity:	29.87%
Query Match:	12.26%
DB:	15
	Gaps:
	16
	270%
	Length:
	Matches:
	Conservative:
	Mismatches:
	Indels:
	110

US-09-864-291-5 (1-313) X US-10-156-761-3543 (1-2709)

QY 97 G L Y T h r I e e L n a l a l A P r o s i g l y l y g l y T r p o l u g l y e l n a l a v a l P h e l y l e u s e r 116
Db 1786 G G G C G C G T A G C G A C G G C C T G S T A C C G G A C C G A C C G A C C G G T A C C C G G T G G C G 1722

```
Oy 117 ---PheArgLySGlyValAlIleGluPheAlaGluLeuMetVal-----Lys 131
Db 1726 GATACCCGTAAGCCCGGGGGCTGCTGACCGGGCATCGGCGGATCGGCTGACCGG 1667
Oy 132 A1a1aSerAla1a1a1aArgLyIleProLeuGlySerValAlaGlyTrpPheAlaPThr 151
Db 1666 GCATCGGCTGGGGGGCCCATGCGCTGCTCCGCTGGGGGGGGGGCCAGGCC--TGGGGACA 1608
Oy 152 SerGlyLeuTrpIleIleThrValProGlyAlaAlaValCysSerSerGlnThrProCys 171
Db 1607 CCGGAGGCGTAACCGGACAGCATGCGCCGAGCA-----CCGGAGACACCCGGC 1560
Oy 172 ProAlaTrpProIleValIleTrpGlyProPro----- 182
Db 1559 GGGGGACAGCCG-----GACCGCCGACCGGGGGACCGGCGACGATCGTCTG 1512
Oy 183 -----ProProGlyTrpThrValGlnProGlyGluTrpGlyThrProProGlu 198
Db 1511 GCGTGTGACACACACCGGGGGCCACCGGGGGCTCCAGAGACACCGGGTGGCGCC----- 1458
Oy 199 GlyTrpGlyAlaGlnProGlyGlyTrpGlyAlaProPrometGlyTrpGlyAlaPro--- 217
Db 1457 ---GGAGGCAATGCCGGGGCACCGGGGGCACCCGGCGGGGAGCGGGCGCGCCCGG 1404
Oy 218 -----ProValGlyTrpGly 222
Db 1403 CCGCCACCGAGGGACCGGCGAGCATGTGTGCGCGGTATGATGACGCGACCGGGA--GCG 1347
Oy 223 ValProProGlyGlyTrpGlyValPro----- 231
Db 1346 GTACCGCCCGGAGTTCGGGGCACCGGGCGGCGGGAGACACCGGCTGACCGGGAGACA 1287
Oy 232 -----ProGlyGlyTrpGlyValPro-----ProGly 240
Db 1286 CCGGGCGGGCCCGGAGGTTCCGGAGCGCCCGAGCGCCGGGCGGACCGGG 1227
Oy 241 GlyTrpGlyAlaPro-----ProGlyGlyTrpGly----- 250
Db 1226 GCGTGGGGCGGCGCGGGCCGGCACCGGGCCCGCCGGGTCCGGAGACATCGTGGCCGCG 1167
Oy 251 -----ValProProGlyGlyTrpGlyAlaProProGlyGlyTrpGlyAlaProProAla 268
Db 1166 TGAATGACACCCCGGGC-----GCGTACCGCCGAGGGGTTCGGCGTACCGGGCCGCG 1113
Oy 269 -----GlyTrpGlyAla----- 272
Db 1112 CTCGGCGGATTCGGGCGCGGAGCGCTCGGGCCCGTCCGGACCGAGTTGGAGAGCGAGC 1053
Oy 273 -----ProProAlaGlyAlaGluAlaLeuProProAlaTrpGlyAlaPro 287
Db 1052 TCGGTGGACATGACCCCGCCCGGGC-----GTACCCGAGACACCGGGCCCG 1005
Oy 288 SerAlaGlyAlaThrAla-AlaSerHisArgSerMetThrAlaGlnGlnGlu 304
Db 1004 GACGCGGAGGAGCATGAGCACCGGGCGCACCCGGCGTTCCAGGGGACACCGGAA 953
```

Search completed: December 16, 2003, 20:57:12
Job time : 337.646 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 16, 2003, 17:31:55 | Search time 2021.93 Seconds

(without alignments)
3762.388 Million cell updates/sec

Title: US-09-864-291-5
Sequence: 1 MAVNQSHTRRGALIPSG.....ASHRMTAQGTSLPTSSS 313

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USFO_spool_p/US09864291/runat.15122003.160859.23318/app.query.fasta.1.846
-DB=BST -QFMT=faaap -SUFFIX=iret -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCLIGN=200 -THR SCORE=Dct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=ext -HRAPIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09864291.QCEN.1.1.3724 @runat.15122003.160859.23318 -NCPU=6 -ICPU=3
-NO MAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXt=7

Database :
EST:
1: em_estda:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vic:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vir:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	983	57.5	1079	12	BM564167
2	935.5	54.7	1414	11	AK015863
3	730.5	42.7	717	14	BY715414
4	653.5	38.2	775	10	BQ722815
5	616	36.0	663	10	BG699398
6	578.5	33.9	932	13	BU107273
7	552	32.3	868	14	CD255288
8	552	32.3	931	13	BU914288
9	540.5	31.6	723	9	AJ452591
10	539.5	31.6	700	13	BU284977
11	536	31.4	636	12	BM490729
12	536	31.4	862	13	BQ735286
13	528.5	30.9	892	14	CD301781
14	517	30.3	719	14	CB593075
15	513	30.0	655	9	AL8898723
16	513	30.0	661	9	AL972453
17	513	30.0	663	9	AL896682
18	506	29.6	647	9	AL849651
19	505	29.5	663	9	AL881810
20	504	29.5	816	13	BU916890
21	500	29.3	568	12	B1390687
22	495	29.0	607	12	BM190975
23	483	28.3	560	13	BU103740
24	481.5	28.1	1165	12	BM923449
25	480.5	28.1	594	12	BM190972
26	480	28.1	735	14	CA364472
27	478.5	28.0	635	9	AL889954
28	476	27.9	613	9	AL889954
29	475	27.8	611	9	AL849211
30	475	27.8	974	13	BQ876783
31	475	27.8	1027	12	BM921835
32	474.5	27.8	970	12	BM424229
33	474.5	27.8	992	12	BM917483
34	472.5	27.6	1024	12	BM451662
35	471.5	27.6	948	13	BU848882
36	471	27.6	865	13	BQ427510
37	471	27.6	1018	12	BM470322
38	470.5	27.5	1068	13	BQ277733
39	470	27.5	914	13	BQ710173
40	469.5	27.5	896	13	BQ212538
41	469.5	27.5	951	14	CD515402
42	468.5	27.4	920	13	BU165608
43	468.5	27.4	1013	12	BM548934
44	468.5	27.4	1042	12	BM479536
45	468	27.4	726	14	CB591559

ALIGNMENTS

RESULT 1
LOCUS BM564167
DEFINITION AGENCOURT_6560149 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:57442129
5', mRNA sequence.
ACCESSION BM564167
VERSION BM564167.1 GI:198811738
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1079)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: L1M1275 row: m column: 10
 High quality sequence scop: 726.

FEATURES
 source location/Qualifiers
 1.1079
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5742129"
 /tissue_type="medulla"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_119"
 /note="Organ: brain; Vector: pCMV-SPORT6, Site_1: NotI; Site_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH MGC Library."

BASE COUNT 267 a 297 c 268 g 247 t
 ORIGIN

Alignment Scores:
 Pred. No.: 2.15e-59 Length: 1079
 Score: 983.00 Matches: 194
 Percent Similarity: 72.56% Conservative: 36
 Best Local Similarity: 61.20% Mismatches: 71
 Query Match: 57.52% Indels: 17
 DB: 12 Gaps: 4

US-09-864-291-5 (1-313) x BMS64167 (1-1079)

QY 2 AAlaValaenGlnSerHisThrGluSerArgGlyAlaLeuIleProSerGlyGluSer 21
 DB 50 GCGGTGAATCAGAGCCACCGAGAACCCCGGAGCCCTCATCCCTTAACGGTGAAGT 109
 QY 22 ValLeuLyGlnCyGluAspValAspLeuCyGpHeLeuGlnLyProValGluSerTy 41
 DB 110 CTTCTGAACGGCTCCGAAATGTGAGCTCTCTCTCCACAGCATCAGAGGCTCAAT 169
 QY 42 LeuPheAsnGlyThyLyLeysGlyThyLeuPheLeuThySerTyArgValAlaPheVal 61
 DB 170 GCTTAGGGTGAAGAACAGAGAACATGTTTTCACATTCATACCGGTGAATTTTCA 229
 QY 62 ThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMetSer 81
 DB 230 ACTTATGCTCATCATGATATCCATGTTGTCTTTATATCATTCATTTGATCTGATGACG 289
 QY 82 AspCyThrIleGlnGlnProIlePheAlaProAsnTyrlleLyGlyThyrlleGlnAla 101
 DB 290 AACCTCACTGTGAACACAGATATTTGCTGCAACTTATTAAGGAATCTATTCAGGCA 349
 QY 102 AlaProGlyGlyGlyTyrPgluGlyGlnAlaValPheLyLeuSerPheArgLyGlyGly 121
 DB 350 GCTCCATATGGTGGCGTGGAGAGCAAGCTTATTAATTAATGCTTTCAGAAATGAGGT 409
 QY 122 AlaIleGluPheAlaGlnLeuMetValLyAlaAlaAsnAlaAlaAlaArgGlyIlePro 141
 DB 410 GCCATTGAATTTGCCAGTGAATGTGAAGGAGTGGCTGTGCTGCCGCCGAGAAATTTCCA 469

QY 142 LeuGlySerValaenTyTrpPheAspThrSerGlyLeuTyrlleIleThrValProGly 161
 DB 470 CTTAGAACCTTAATATGACTGGTTCAGCTCATAGGAAATTTATTAATACGCGGAGAGG 529
 QY 162 AAlaValaValSerSerGlnThrProCySerProAlaTyProIleValIleTyrlPro 181
 DB 530 --AATATGTCACCTCCACCAATGCTGTG-----TCAGTATGTTCTATGAGGCC 577
 QY 182 ProProGlyGlyTyTrpValGlnProGlyGlyTyTrpProGlyGlyTyTrpGly 201
 DB 578 CCACTGAGAGA-----TATGAGGCCCACTCCCGGATACGGA 616
 QY 202 AlaGlnProGlyGlyTyTrpValaProProMetGlyTyTrpValaProProValaGly 221
 DB 617 GCCCACCCTGAGATATGAGGCCCAACCCGTAGAAATGAAGCCCGCTGTGGATAC 676
 QY 222 GlyValaProProGlyGlyTyTrpValaProProGlyGlyTyTrpValaProProGly 241
 DB 677 AGAGCTTCACCTGTCGATATGAGGCCCACTCTTGATATACGAGGCCCACTGACAGA 736
 QY 242 TyrlGlyAla-ProProGlyGlyTyTrpValaProProGlyGlyTyTrpValaProProGly 261
 DB 737 TATGAGCCCTCACTGATATGAGGCCCACTCTTGATATGAGGCCCACTCT 796
 QY 261 GylTyTrpValaProProAlaGlyTyTrpValaProProAlaGlyAsnGlnAlaLeuPr 281
 DB 797 CGATATATGAGCCCACTCTCGATATGAGGCCCACTCTGAGAAATGAAGCCCGGCC 856
 QY 281 CProAlaTyTrpValaProProAlaGlyAsnGlnAla-----AlaSerHisAr 297
 DB 857 TGGCGGATACAGAGCTTCACCTGTCGATATGAGGCCCAAGGCTTCGAATTCAGC-AG 915
 QY 297 gSerMetThrAlaGlnGlnGlnThrSerLeuProThrThrSerSerSer 313
 DB 916 GCCAGGCTCCCTGAACAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 964

RESULT 2
 AKO15863
 LOCUS
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930521123 product:hypothetical protein, full insert sequence.

ACCESSION AKO15863
 VERSION AKO15863.1 GI:12854363
 KEYWORDS HTC, CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

REFERENCE
 1 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)

REFERENCE
 2
 10349636
 TITLE
 JOURNAL
 PUBMED
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

REFERENCE
 3
 20499374
 11042159
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

Db	Accession	Version	KeyWords	Source	Organism	Reference	Authors
Dp	529						
Dp	181						
Dp	577						
Dp	201						
Dp	637						
Dp	217						
Dp	697						
Dp	218						
Dp	757						
Dp	235						
Dp	817						
Dp	248						
Dp	877						
Dp	268						
Dp	937						
Dp	283						
Dp	997						
Dp	301						
Dp	1057						
Dp	BY715414						
Dp	LOCUS						
Dp	DEFINITION						
Dp	ACCESSION						
Dp	VERSION						
Dp	KEYWORDS						
Dp	SOURCE						
Dp	ORGANISM						
Dp	REFERENCE						
Dp	AUTHORS						

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

22354683
12466851

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shohiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.go.jp,
url:http://genome.gsc.riken.go.jp/
Aachit,U., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
S., Hashizume,M., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Komio
H., Koye,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
Waki,K., Watanabe,A., Muramatsu,M. and Hayashizaki,Y. Direct
Submission

Computational Analysis of Full-length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1611-1630 (2000)
RIKEN Integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

location/Qualifiers

1..717

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="4930521123"

/sex="male"

/tissue_type="testis"

/dev_stage="adult"

/lab_host="SOLR"

/clone_1b="RIKEN full-length enriched, adult male testis"

/note="Site 1: XhoI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5']
GAGGAGAGAGAGCGCGCATTAATTCGAGGTAATTAATTAATCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites. "

161 a 193 c 198 g 165 t

Alignment Scores:

Pred. No.: 6,896-42 Length: 717
 Score: 730.50 Matches: 145
 Percent Similarity: 64.45% Conservative: 20
 Best Local Similarity: 56.64% Mismatches: 54
 Query Match: 42.74% Indels: 37
 DB: 14 Gaps: 4

US-09-864-291-5 (1-313) x BY715414 (1-717)

```

Qy 1 MetAlaValAndInSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 55 ATGCACTGCAACGACCAACATACAGTGGACCGCGTGGCGCCATCCCTCATGGTGA 114
Qy 21 SerValLeuYsgInCyGluAspValAspLeuCySphLeuGlnIlyProValGluSer 40
Db 115 AGTCTCTTGAAGAATGTCACAAAGTGAACCTCTCCCTCCCGAGTCAACCCGAGCTCC 174
Qy 41 TytleupheAnglyThrIlySgIlyThrIleupheLeuThrSerTyTzArgValIlePhe 60
Db 175 AACCTCTTGAAGTGAAGAAGAGGAGGCTCTGTTCTCACTCATACCGAGTGAATTT 234
Qy 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
Db 235 GTGACTTCGCGTGCAGACACATCCATGTTTCTTTACAGATCCATTCATCATATG 294
Qy 81 SerAspCyThrIleGluGlnProIlePheAlaProAsnTyTleYsgIlyThrIleGln 100
Db 295 AATAAGTCACTGTTGAACAACCAATCTTGTGCAAACTCATTTAAAGGAGCAATTCAG 354
Qy 101 AlaAlaProGlyGlyGlyTPGglGlyGlnAlaValPheIlySleuSerPheArgIly 120
Db 355 GCAGCTCCAGAGCGGTGGGAGGAGTTCGCTTCTTAAATGAATCTTCGGAAGA 414
Qy 121 G1ValIleGluPheAlaGlnLeuMetValIlyAlaIleAlaIleAlaIleAlaIle 140
Db 415 GGTCTATTCAGCTTGTCCAGTGTATGCGCAAGCTCTGCGGCTGCCAGAGT 474
Qy 141 ProLeuGlySerValAsnTyTzPheAspThrSerGlyLeuTyTleIleThrValPro 160
Db 475 CCACTGACAGATGACAGCTCTGATGAGGCGCTCAAGAAATTTATGATCACT 528
Qy 161 G1ValAlaValCySerSerGlnThrProCySproAlaTyTzProIleValIleTyGly 180
Db 528 ----- 528
Qy 181 ProProProGlyTyTzThrValGlnProGlyGlyTyTzProProGlyGly 199
Db 529 -----GGGACAGAAACATGATATGCCCCACAGCA 558
Qy 200 TyTzIlyAlaGlnProGlyGlyTyTzIlyAlaProProMetGlyTyTzIlyAlaProProVal 219
Db 559 TACCAAGTT-----GCTATGAGAGCCCACTCGGAGATATGAGGCTCAACCGTG 609
Qy 220 G1TyTzIlyValProProGlyGlyTyTzIlyValProProGlyGlyTyTzIlyValProPro 235
Db 610 GGAATACGAGATCCATCTGACAGGATATGAGACCCCACTCGGAGTATGAGAGCCGCC 669
Qy 240 G1GlyTyTzIlyAlaProProGlyGlyTyTzIlyValProProGlyGly 254
Db 670 GTGGATATGATACCCCACTCTCCGAGATGATGCTGCTCTGGA 717

```

RESULT 4

BG722815

LOCUS BG722815 775 bp mRNA linear EST 08-MAY-2001

DEFINITION 602695278P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4827427 5',

mRNA sequence.

ACCESSION BG722815

VERSION BG722815.1 GI:14002002

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Rutheria; Primates; Catarrhini; Homnidae; Homo.
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Straube, Ph.D.
 Email: gsabbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovics, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
 Toshiyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
<http://image.llnl.gov>
 Plate: LIML0743 row: 1 column: 20
 High quality sequence stop: 651.
 Location/Qualifiers

FEATURES

source

1. 775
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4827427"
 /lab_host="DH10B"
 /clone_1lb="NIH_MGC_97"
 /note="Organ: testis; Vector: pBluescript (modified
 pBluescript KS+); Site: 1: BamHI, Site 2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTATTTTATTTTATTT-3',
 size-selected for average insert size 2.2 kb and
 normalized to ROT 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC library."
 ORIGIN 191 a 212 c 193 g 179 t

BASE COUNT

Alignment Scores:

Pred. No.: 1,876-36 Length: 775
 Score: 653.50 Matches: 144
 Percent Similarity: 71.37% Conservative: 28
 Best Local Similarity: 59.75% Mismatches: 48
 Query Match: 38.24% Indels: 23
 DB: 10 Gaps: 4

US-09-864-291-5 (1-313) x BG722815 (1-775)

```

Qy 1 MetAlaValAndInSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
Db 54 ATGCACTGCAACGACCAACATACAGTGGACCGCGTGGCGCCATCCCTCATGGTGA 113
Qy 21 SerValLeuYsgInCyGluAspValAspLeuCySphLeuGlnIlyProValGluSer 40
Db 114 AGTCTCTTGAAGAATGTCACAAAGTGAACCTCTCCCTCCAGATCCAGAAAGCTCA 173
Qy 41 TytleupheAnglyThrIlySgIlyThrIleupheLeuThrSerTyTzArgValIlePhe 60
Db 174 AACCTCTTGAAGTGAAGAAGAGGAGGCTCTGTTCTCACTCATACCGAGTGAATTT 233
Qy 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
Db 234 ATAACCTGATGCTCAACAGTGAATCCCATGTTTCTTTATGATGATGATGATGATG 293
Qy 81 SerAspCyThrIleGluGlnProIlePheAlaProAsnTyTleYsgIlyThrIleGln 100
Db 294 ACAACCTGATGCTGTAACAACCAATGTTGCTGCAAACTCTTTAAGGAACTATTGAG 353
Qy 101 AlaAlaProGlyGlyGlyTPGglGlyGlnAlaValPheIlySleuSerPheArgIly 120
Db 354 GCAGCTCCATATGATGCTGGAAGAGCAAGTACTTATTAATTAATTAATTAATTAATGA 413
Qy 121 G1ValIleGluPheAlaGlnLeuMetValIlyAlaIleAlaIleAlaIleAlaIle 140

```

Db 414 GGTGCCATTGAATTTGCCCAATTGATGGTAAAGCTGCTCTGCTGCTGCCGAGAAATT 473
 Qy 141 ProLeuGlySerValAsn-TyTrpPheAspThrseryLeuTyrlleIleThrValPr 160
 Db 474 CCACCTTAGAACCTTAACGACTGGTTCAGCTCTATGGGAATTATGTAATTAACGGGGA 533
 Qy 160 GGIYAlaValCysSerSerGlnThrProCysProAlaTyrrProIleValIleTyrgl 180
 Db 534 AGGGC--AATATGTCACCTCCACAGATGCTTGT-----TCAGTTATGCTATAG 581
 Qy 180 YrProProProGlyIYrThrValGlnProGlyGluTyrrThrProProGluGlyT 200
 Db 562 ACCCCCACTGAGGAGAT-----GAGCCCACTCCCGGAT 620
 Qy 200 YrGIYAlaGlnProGlyIYrGIYAlaProProMetGlyTyrrGIYAlaProProValG 220
 Db 621 ACGGAGCCCACT-GCAGGATATGAGCCCAACCCGTAGAAA-GAGCGCGCGCTGCTG 678
 Qy 220 IYTr-----GlyValProProGlyIYrGIYValPro 231
 Db 679 GATATACAGAGCTCAACCTGTCGACAAATGGAAGCCCACTCTCGGCAATACGCGAGCCC 737
 RESULT 5
 LOCUS BG699398 663 bp mRNA linear EST 07-MAY-2001
 DEFINITION 602679081P1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811804 5',
 mRNA sequence.
 ACCESSION BG699398
 VERSION BG699398.1 GI:13967653
 KEYWORDS EST.
 ORGANISM Homo sapiens (human)
 SOURCE
 REFERENCE 1 NIH-MGC http://mgs.nci.nih.gov/
 AUTHORS 1 (bases 1 to 663)
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgarbs-remail.nih.gov
 Tissue procurement: Mikhail Palkovits, M.D., Ph.D.
 cDNA library preparation: Michael J. Brownstein (NHLBI), Shitaki
 Toshitaki and Piero Carninci (RIKEN)
 cDNA library arrayed by: The I.M.A.G.E. Consortium (ILMM)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILMM at:
 http://image.llnl.gov
 Plate: LAM10703 row: a column: 21
 High quality sequence stop: 660.
 FEATURES
 source
 1. 663
 location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4811804"
 /issue_type="hippocampus"
 /lab_host="DH10B"
 /clone_11b="NIH_MGC_95"
 /note="Organ: Brain; Vector: pBluescript (modified
 pBluescript KS+); Site 1: BamHI, Site 2: SalI-XhoI (gtcag
 ; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
 size selected for average insert size 2.5 kb and
 normalized to RGT 5. This is a primary library enriched
 for full-length clones and constructed using the
 cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIMH/NHLBI, National
 Institutes of Health). Note: this is a NIH-MGC library."
 BASE COUNT 168 a 158 c 169 g 164 t 4 others
 ORIGIN
 Alignment Scores: 6.71e-34 Length: 663
 Pred. No.:

Score: 616.00 Matches: 134
 Percent Similarity: 74.09% Conservative: 29
 Best Local Similarity: 60.91% Mismatches: 46
 Query Match: 36.04% Indels: 14
 DB: 10 Gaps: 3
 US-09-864-291-5 (1-313) x BG699398 (1-663)
 Qy 1 MetAlaValangInserHlethrgLuserArgTgIYAlaLeuIleProSerGlyGlu 20
 Db 37 ATGGCGGATGATGAGACCAACAGAGAACCGCGCGAGCCCTCATCCTAACGGTGA 96
 Qy 21 SerValIleuysGlnCysGluAspValAspLeuCySphLeuGlnIlyProValGluSer 40
 Db 97 AGTCTCTTGAAGCCGTCCTCCGAATGTGAGGCTCTCTCCACAGCATCAGAGAGCTCA 156
 Qy 41 TyrlleuPheangIYrThrValYsGlyThrlleuPheuthrserTyrrValValPhe 60
 Db 157 AATGCTTATGTGTAGAAAGAACAGAACATG-TTCTCACTTCATACCGGATATTTC 215
 Qy 61 ValThrseryIleuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db 216 ATAACTTCATGCTGCATCAGATGATCCATGATGCTTTATGATGACATTTGATCTGATG 275
 Qy 81 SerAspCythrIleGluGlnProIlePheAlaProAlaTyrrIleYsGlyThrlleGln 100
 Db 276 ACGAACCTCACTGTGAAACCAACAGATTGCTCAACTTCAATTAAGGAACTAATTCAG 335
 Qy 101 AlaAlaProGlyIYrGIYTrpGIYGIYAlaValPheYsLeuSerPheArgIYrGIY 120
 Db 336 GCAAGCTCATATGTGTGTGGAGAGACAGACTTCTTAAATTAATGCTTCAGAAATGGA 395
 Qy 121 GIYAlaIleGluPheAlaGlnLeuMetValYsAlaIleAlaIleAlaIleArgIYrIle 140
 Db 396 GATGCATTAATGATTTGGCCAGTTGATGTGAAGACCTCTCTGCTGCGCCGAGAAATT 455
 Qy 141 ProLeuGlySerValAsnTyrrThrPheAspThrseryLeuTyrlleIleThrValPro 160
 Db 456 CCACCTTAGAACCTTAATTAATGACTGCTGTGAGCTTATGTAATTAACGTGGGAA 515
 Qy 161 GIYAlaIleValCysSerSerGlnThrProCysProAlaTyrrProIleValIleTyrgl 180
 Db 516 GGG--AATATGTCACCTCCACAGATGCTTGT-----TCAGTTATGCTATAGGA 563
 Qy 181 ProProProProGlyIYrThrValGlnProGlyIYrGIYThrProProGluGlyTyrr 200
 Db 564 GCCCACTGAGAG-----TATGAGCCCACT-CCCGGATAC 601
 Qy 201 GIYAlaGlnProGlyIYrGIYAlaProProMetGlyTyrrGIYAlaProProValGly 220
 Db 602 GAGCGCCC-ACATGAGATATGAGAGCCCAACCGTAGAATGAAGCGCGCTGTGGGA 660
 RESULT 6
 LOCUS BU107273 932 bp mRNA linear EST 25-NOV-2002
 DEFINITION 603111954P1 CSBCHU12 Gallus gallus cDNA clone CHEST60K13 5', mRNA
 sequence.
 ACCESSION BU107273
 VERSION BU107273.1 GI:25309754
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 REFERENCE Boardman, P.B., Samz-Bazquero, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.,
 A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PMID 12445392
 COMMENT Contact: Simon Hubbard

Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK
Tel: 0161208930
Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

FEATURES

SOURCE

1. 932
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CBST60k13"
/dev_stage="35"
/lab_host="DH10B"
/clone_id="CSBOCHL12"
/note="Organ: heart; Vector: pBluescript II KS(+); Site 1:
Scot1; Site 2: Not1; Modification of pBluescript II KS(+)
[Stratagene] vector to accommodate cDNA produced with the
T-titrimed protocol (construction of uni-directionally
cloned cDNA libraries from messenger RNA for improved 3'
end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
, 624). Cut pBluescript II KS(+) with NotI and Scot1.
Ligate in double stranded adaptor containing BglI and
BamHI sites (5'ggcgccgcgcagcccgagccgaaaaaag)
[5'aattcttttcgagatccgggctcgcagc]"

BASE COUNT 215 a 273 c 233 g 211 t

ORIGIN

Alignment Scores:

Pred. No.: 4,12e-31 Length: 932
Score: 578.50 Matches: 131
Percent Similarity: 49.21% Conservative: 24
Best Local Similarity: 41.59% Mismatches: 81
Query Match: 33.85% Indels: 80
DB: 13 Gaps: 8

US-09-864-291-5 (1-313) x BU107273 (1-932)

QY 1 MetAlaValAngInSerHisThrGluSerArgAlaLeuIleProSerGlyIu 20
Db 18 ATGGCGCTCATATAGCACCACTCGAAG---GAAGCGGTGTCCTCTCCCAAGCCGAG 74
QY 21 SerValLeuYngInYngValAspLeuCySpheLeuGlnYpProValGluSer 40
Db 75 AGCATTTCTCAAGATGTAAGATGTGAGCTCTCCAGTGCATGACGAGCGCAAGCTT 134
QY 41 TyrlpheAngInYThrIlyGlyGlyThrlpheLeuThrSerYTrAqValValPhe 60
Db 135 GAGGCTTCAGAGCACCAAGAGGAGATGCTGTATCTCACCCCATTCAGAGATGATCTTC 194
QY 61 ValThrSerHisLeuValAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
Db 195 GTTTCAGAAAGGCG---AAGATCTCTATGCTGCTTTTATGATGCCGTTTATTTGATG 248
QY 81 SerAspCyTrHrIlegInGInProIlePheAlaProAsnYTrIleYngIYThrIlegIn 100
Db 249 AAAGGGGTCTATATGACAGCCCTTTCTCTGCTAATTCATCAAGAGCAAGATTGAG 308
QY 101 AlaAlaProGlyGlyIYTrpGInGlyGlnAlaValPheIlyLeuSerPheAsyGlyGly 120
Db 309 GCTGAGGACAGAGGTGTGCTGGAAGGACAGGAACTTTAACTGACATTCACAGCGGA 368
QY 121 GlyValAlIlegIupheAlaGInLeuMetValIyAlaIAlaSerAlaAlaIArgGlyIle 140
Db 369 GGAAGCATTCAGATTGACACACTCATGTTCAAACTGCTCTTACGTCTTCACAGTGAATT 428
QY 141 ProLeuGlySerValAsnYTrIYTrpPheAspThrSerGlyLeuYTrIleIleThValPro 160
Db 429 CCTCTC----- 434
QY 161 GlyAlaAlaValCySerSerGInThrProCySpProAlaYTrProIleValIleYTrGly 180

Db 434 ----- 434
QY 181 ProProProProGlyYTrYThrValGInProGlyGInYTrYThrProProGlyYTr 200
Db 435 ---CAAGCCCCCTGGCTAT-----GGATAC 455
QY 201 GlyAlaGInProGlyYTrYTrYValAlaProProMetGlyYTrYValAlaProProValGly 220
Db 456 ACAACCTGACTGAGAGGTATGACACTGTCCCA-----CCTGCTCAGAGAGG 503
QY 221 YTrGlyValProProGlyGlyYTrYTrYValAlaProProGlyYTrYTrYValAlaProProGly 240
Db 504 TATGACACTGCTCAGAGGTATGACACTGTCTCAGAGGTATGACACTGCTCTCAGGA 563
QY 241 GlyYTrGlyValAlaProProGlyYTrYTrYValAlaProProGlyY----- 254
Db 564 GGGTATGACACTGCTCCGGGGGCTATGCTCTCCACCC-ACTCGAATATGCTTAT 622
QY 255 GlyYTrGlyValAlaProPro---GlyGlyYTrGlyYValAlaProProAlaGlyYTrGlyYValAlaPro 273
Db 623 CCTTATGACACCACTCCGATGAAATGCTTACGACCTGCTCCAGCCCATGGGGTATCCG 682
QY 274 ProAlaGlyYAngInuAlaLeu----- 280
Db 683 TATGCCAGATTCAGATTATTTACCACTCCCTCCACATGACATTCACATATGTGATC 742
QY 281 ---ProProAlaYTrGlyValAlaProSerAlaGlyYAsnThrAlaAla 294
Db 743 CTCCTCTCCCTCATCTCCGGGCGCATCTCTCGAAGACCTCAGCC 787

RESULT 7
CD255258
LOCUS
DEFINITION
AGENE COUNT 14156849 NICHD XGC Brn1 Xenopus laevis cDNA clone
IMAGE:6954461 5', mRNA sequence.
CD255258
VERSION
CD255258.1 GI:31015724
EST.
Xenopus laevis (African clawed frog)
SOURCE
Xenopus laevis
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 868)
NIH-MGC http://mgc.nci.nih.gov/
TITL
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
COMMENT
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cga@pds-remail.nih.gov
Tissue Procurement:
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov.
Plate: LNA14578 row: g column: 04
High quality sequence stop: 673.
Location/Qualifiers

FEATURES

SOURCE

1. 868
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:6954461"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/clone_id="NICHD XGC Brn1"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI,
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.5 kb. Constructed by Life Technologies. Note: This is a Xenopus Gene Collection (XGC) library.

BASE COUNT 241 a 244 c 189 g 193 t 1 others

Alignment Scores:

Pred. No.: 2,756-29 Length: 868
Score: 552.00 Matches: 131
Percent Similarity: 53.44% Conservative: 32
Best Local Similarity: 42.95% Mismatches: 88
Query Match: 32.30% Indels: 54
Gaps: 13

US-09-864-291-5 (1-313) x CD255258 (1-868)

1 MetAlaValAnGlnSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
40 ATGTCACTTAATTAAGAACCACTCAAGGGG--GCCGGATCATGTGTACGACAGGGAG 96
21 SerValLeuLeuGlnCysGlnuAspValAspLeuCysPheLeuGlnuAspProValGluSer 40
97 AGTATTCTCAGACAGTGTAAAGATGTGGAACCTCTCTCAGGATATGTCTCAGCAATCA 156
41 TyrLeuPheAnGlyThrIleYleGlyThrIlePheLeuThrSerTyrArgValPhe 60
157 GAAGCATTCAAAGGACAAAGAAAGAGATGCTATATCTCACCGCTACAGGGGTATCTTT 216
61 ValThrSerHisLeuValAspAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
217 CTAAACCAAGG-----AAAGACCCCATGATGCTCTTCAATATATGCTATCACTGATG 270
81 SerAspCysThrIleGlnIleProIlePheAlaProAsnTyrIleGlyThrIleGln 100
271 AAAGATGCTCATTAAGAACCACTCTCTCTGCAACTACATCAAGGAAACCATCACT 330
101 AlaAlaProGlyGlyIleTyrGlnuGlnAlaValPheYleLeuSerPheArgGly 120
331 GCTGAACCTGAGATGCTGGAGGGTAAACCTTTTAACTTAACTTAACTTAACTGGA 390
121 GlyAlaIleGlnPheAlaGlnLeuMetValYleAlaAlaSerAlaAlaAlaArgGlyLe 140
391 GGAGCTATTGAATTTGACAAATCATGTTCAAAATGGCACTTGGCTCCAGACCCCT 450
141 ProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleThrValPro 160
451 CCT-----GTTCT 459
161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGly 180
460 CATGCTGCATAT-----GATATACACTCTCTGCTGAGGA-----GGA 495
181 ProProProProGlyTyrThrValGlnProGlyGlyTyrGlyThrProProGlyGlyTyr 200
496 TATGACCAAGGGGGTATCCCTCCAGCTCCAGGAATATATACACCCCAACCAACATCA 555
201 GlyAlaGlnProGlyGlyTyrGlyAlaProPheMet-----GlyTyrGlyAlaProVal 219
556 GGCCCTTATCA-----TATGACCCCTCTGCTATGATGATATGACCACTCCGAT 609
220 GlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyValProPro 239
610 GCTATGGGATACCG-----TATGCTCCGCAACAGGA-----CCT 645
240 GlyGlyTyrGlyAlaProProGlyGlyTyrGlyValProProGlyGlyTyrGlyAlaPro 259
646 GGCATGTATCAACCCCTCTCTGATATGATCCATT-----TACATGCTCTCT 693
260 ProGlyGlyTyrGlyAlaProProAlaGlyTyrGlyAlaProProAlaGlyAlaGlyAla 279
694 CCAACCACTTATCTCGAGCCCG-----TATATATGACCCCACTATATATGA----- 741
280 LeuProProAlaTyrGlyAlaProSerAlaGly-----AsnThrAla 293

Db 742 ---ACCCAGCACTTCAAGCTCTCTGATTCATGATATCTGATATCTGAGGAGCAG 798
Qy 294 AlaSerHisArgSer 298
Db 799 CAAGCCACAGAGAC 813

RESULT 8
BU914288 931 bp mRNA linear EST 17-OCT-2002
LOCUS AGENCOURT 10492155 NICHD XGC OOI Xenopus laevis cDNA clone
DEFINITION IMAGE:6640777 5', mRNA sequence.
ACCESSION BU914288
VERSION BU914288.1 GI:24096202
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis

REFERENCE
AUTHORS NCI-CGAP
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
JOURNAL
COMMENT Unpublished
Contact: Robert Strauberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LIML4223 row: 1 column: 01
High quality sequence stop: 654.

FEATURES

source
1..931
Location/Qualifiers
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8155"
/clone="IMAGE:6640777"
/feature="occytes"
/lab_host="DH10B (phage-resistant)"
/clone_id="NICHD XGC OOI"
/note="Vector: pCMV-SPORT6, Site 1: NotI, Site 2: SalI;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.2 kb. Constructed by Life Technologies."
BASE COUNT 254 a 256 c 201 g 220 t
ORIGIN

Alignment Scores:

Pred. No.: 2,966-29 Length: 931
Score: 552.00 Matches: 133
Percent Similarity: 55.22% Conservative: 31
Best Local Similarity: 44.78% Mismatches: 91
Query Match: 32.30% Indels: 42
Gaps: 12

US-09-864-291-5 (1-313) x BU914288 (1-931)

1 MetAlaValAnGlnSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
62 ATGTCACTTAATTAAGAACCACTCAAGGGG--GCCGGATCATGTGTACGACAGGGAG 118
21 SerValLeuLeuGlnCysGlnuAspValAspLeuCysPheLeuGlnuAspProValGluSer 40
119 AGTATTCTCAGACAGTGTAAAGATGTGGAACCTCTCTCAGGATATGTCTCAGCAATCA 178
41 TyrLeuPheAnGlyThrIleYleGlyThrIlePheLeuThrSerTyrArgValPhe 60
179 GAAGCATTCAGAGGACAAAGAAAGAGATGCTATATCTCACCGCTTACAGGGGTATCTTC 238

QY 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 239 CTAAACAAGGG-----AAGATCCCATATGCTTTAATAGCCATTCTACATATG 292
 QY 81 SerAspCysThrIleGluGlnProIlePheAlaProAntyTyleGlyThrIleGln 100
 DB 293 AAGAGATGCTCCATAGAACAGCCAGGTCTCTCCCAATCAATCAAGGAAACCATAGT 352
 QY 101 AlAlaAlaProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 120
 DB 353 GCTGAACCTGAGGTGCTGAGGAGGTCAAGCTCTTTTAAATTAATTAATTAATTAATG 412
 QY 121 GYAlaIleGluPheAlaGlnLeuMetValIleValIleAlaIleAlaIleAlaIleAla 140
 DB 413 GGTCTATTGAATTTGACAAATCATGTCTAAATGACAACTTGCTTCCAGAACCCCT 472
 QY 141 ProLeuGlySerValAsnTyrrTrpPheAspThrSerGlyLeuTyrrIleIleThrValPro 160
 DB 473 CCT-----GTTCT 481
 QY 161 GYAlaAlaValCysSerSerGlnThrProCysProAlaTyrrProIleValIleTyrrGly 180
 DB 482 CATGCTGCATAT-----GGATACACACCTGCTCCAGAG-----GGC 517
 QY 181 ProProProProGlyTyrrThrValGlyProGlyGlyTyrrThrProProGlyGlyTyrr 200
 DB 518 TATGCAATAGCGGGATCCCTCCAGCTCCAGAAATATATACACCCCAACCCCAACATCA 577
 QY 201 GYAlaGlnProGlyGlyTyrrGlyAlaProPheMet-----GlyTyrrGlyAlaPro----- 217
 DB 578 GGCCCTTATCCA-----TATGACCCCTGCTATGATGATGATGATGATGATGATGATGAT 631
 QY 218 ProValGlyTyrrGly-----ValProProGlyGlyTyrrGlyValProProGlyGlyTyrrGly 236
 DB 632 CCAATGGAATACCCGTATGCTCCACAGAGAG-----CCGCGCATGATACCA 679
 QY 237 ValProProGlyGly-----TyrrGlyAlaProProGlyGlyTyrrGlyValProPro 253
 DB 680 CCCCCTCTGAATGAATCCCATTTACATGAGCTCTCCACACCTTATCTCGAGACCCCC 739
 QY 254 GYTyrrGlyAlaProProGlyGlyTyrrGlyAlaProProAlaGlyTyrrGlyAlaPro 273
 DB 740 TACAT-----GGAGCCCTTCAATGAAACCCAGCACCTTCTGCTCCACATGATGATG 796
 QY 274 ProAlaGlyAsnGlyAlaLeuProProAlaTyrrGlyAlaProSerAlaGly 290
 DB 797 AATGCTGTATGCTGAGAGTGAAGCAAGCTGCAAAAGCTGCATCCAGTGGC 847
 RESULT 9
 AJ452591 723 bp mRNA linear EST 22-APR-2002
 LOCUS AJ452591 riken1 Gallus gallus cDNA clone 31hr1, mRNA sequence.
 DEFINITION AJ452591
 ACCESSION AJ452591
 VERSION AJ452591.1 GI:20262687
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 TISSUE Erythrocytes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianidae; Gallus.
 REFERENCE 1 (bases 1 to 723)
 AUTHORS Buerstedde, J.M.
 JOURNAL Gallus gallus Bursal Lymphocyte EST
 COMMENT Unpublished
 CONTACT: Buerstedde JM
 Cellular Immunology
 Heinrich-Pette-Institute
 Martinstr. 52, 20251 Hamburg, Germany
 Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
 FEATURES
 1..723
 location/Qualifiers
 1..723
 /organism="Gallus gallus"
 /mol_type="mRNA"

/db_xref="taxon:9031"
 /clone="31hr1"
 /cell_type="Bursal Lymphocyte"
 /dev_stage="2-3 weeks old"
 /clone_lib="riken1"
 /note="CB inbred strain"
 BASE COUNT 161 a 205 c 181 g 172 t 4 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,44e-28 Length: 723
 Score: 540.50 Matches: 126
 Percent Similarity: 52.23% Conservative: 26
 Best Local Similarity: 43.30% Mismatches: 66
 Query Match: 31.63% Indels: 74
 DB: 9 Gaps: 11
 US-09-864-291-5 (1-313) x AJ452591 (1-723)
 QY 1 MetAlaValAsnGlnSerHisThrGluSerArgArgGlyAlaLeuIleProSerGlyGlu 20
 DB 12 ATGGCGCTCATATAGGACCACTGAG-----GAAAGCGGTGTCGTGCTCCCAATGCCAG 68
 QY 21 SerValLeuGlyGlnCysGluAspValAspLeuCysPheLeuGlnTyrrProValGluSer 40
 DB 69 AGCATTTCAAGCAATGTAAAGATGTAGAGCTCTCTTCAATGACATGACAGGCAAGCTT 128
 QY 41 TyrrPheAlaGlyTyrrIleValGlyGlyTyrrLeuPheLeuThrSerTyrrArgValValPhe 60
 DB 129 GAGGCTTCAAAGGACCAAGAGGAAATGCTATCTCAACCCATACAGATGATCTTC 188
 QY 61 ValThrSerHisLeuValAsnAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 189 GTTCTCAAGGCG-----AAGATCTCATGCTGTCTTTATATATATATATATATATAT 242
 QY 81 SerAspCysThrIleGluGlnProIlePheAlaProAntyTyleGlyThrIleGln 100
 DB 243 AAGAGTGTCTTATGACAGCTCTTTCTCTGCTAATTAATCAAAAGACATTCAG 302
 QY 101 AlAlaAlaProGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 120
 DB 303 GCTGAAGGAGAGGTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 362
 QY 121 GYAlaIleGluPheAlaGlnLeuMetValIleValIleAlaIleAlaIleAlaIleAla 140
 DB 363 GAGGCATCATGATTTGACAGCTGATGTTCAAAAGCTGCTCTGCTTCCAGTGAAGTT 422
 QY 141 ProLeuGlySerValAsnTyrrTrpPheAspThrSerGlyLeuTyrrIleIleThrValPro 160
 DB 423 CTTCTC----- 428
 QY 161 GYAlaAlaValCysSerSerGlnThrProCysProAlaTyrrProIleValIleTyrrGly 180
 DB 428 ----- 428
 QY 181 ProProProProGlyTyrrThrValGlnProGlyGlyTyrrThrProProGlyGlyTyrr 200
 DB 429 -----CAAGCCCTGCTAT----- 449
 QY 201 GYAlaGlnProGlyGlyTyrrGlyAlaProProMetGlyTyrrGlyAlaProProValGly 220
 DB 450 ACACTGTACCTGAGAGGTAT-----GCACCT-----GTCCCACTGCT----- 488
 QY 221 TyrrGlyValProProGlyGlyTyrrGlyValProProGlyGlyTyrrGlyValProProGly 240
 DB 489 -----CCAGAGAGGTATGACCTGCTCCAGAGAGGTATGACCTGCTCCGGGG 536
 QY 241 GYTyrrGlyAlaProProGlyGlyTyrrGlyValProProGlyGly-----TyrrGly 257
 DB 537 GGCTAT-----GCTCTCTC-----NCAACCACTCCAAATGATGCTTATCTTATGCA 584
 QY 258 AlaProPro-----GlyGlyTyrrGlyAlaProProAlaGlyTyrrGlyAlaProProAlaGly 276

Db 585 CCACCTCCGATGATGCTACGAGACTGCTCCCAACCCATGAGGAT-CCGTATGCCAG 643
 Qy 277 AanglualeuProProbaletyrglualeuPro 287
 Db 644 MATCCAGATTTTATCCACCCGCTGCCACCA 676
 RESULT 10
 B0284977 700 bp mRNA linear EST 27-NOV-2002
 LOCUS B0284977 CSEGCN54 Gallus gallus CDNA clone CSEST581n20 5', mRNA
 DEFINITION
 ACCESSION B0284977 GI:25734433
 VERSION B0284977.1
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 700)
 Boardman, P.E., Sanz-Bzquez, J., Overton, I.M., Burt, D.M., Bosch, B.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken CDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 2235534
 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 0161206930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 Location/Qualifiers
 1..700
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="Compton line 15T"
 /db_xref="taxon:9031"
 /clone="CSEST581n20"
 /sex="Female"
 /tissue_type="not cerebrum or cerebellum"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_1lb="CSEGCN54"
 /note="Organ: Brain; Vector: pBluescript II KS(+); Site_1:
 EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunt-ended, ligated to NotI adapters, digested with EcoRI
 , size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

BASE COUNT 162 a 196 c 178 g 164 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1.64e-28 Length: 700
 Score: 539.50 Matches: 124
 Percent Similarity: 51.75% Conservative: 24
 Best Local Similarity: 43.36% Mismatches: 67
 Query Match: 31.57% Indels: 72
 DB: 13 Gaps: 9

US-09-864-291-5 (1-313) x B0284977 (1-700)

Qy 1 MetAlaValAangInserHisThrGluSerArgArgValAlaLeuLeuProSerGlyGlu 20
 Db 3 ATGGGCTCAATAGGACCACTCGAAG---GAGGGGCTGTCGTCGCCATGCCAG 59
 Qy 21 SerValLeuValGlnGlyGluValAspValAspLeuGlySerPheLeuGlnValProValGluSer 40
 Db 60 AGCATTCGACGAAATGTAAGATGTGAGCTCTCTCACTGACATGACGAGGAGCTT 119
 Qy 41 TyrLeuPheAangInGlyThrLeuValGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
 Db 120 GAGGCTTCAAGGACCAAGAGGAGATGCTGATCTACACCCATACGATATCTTC 179
 Qy 61 ValThrSerHisLeuValAspAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db 180 GTGTCAAGAGC-----AAGATCTCATGCTGCTTTATGATGACCGTTTATTGGTG 233
 Qy 81 SerAspGlyThrIleGluGlnProIlePheAlaProAlaGlyIleValGlyThrIleGln 100
 Db 224 AAGGGTCTCATATGACGACCTCTTCTCTCATATTCATCAAGAGGACAGATTCAG 293
 Qy 101 AlaAlaProGlyGlyGlyTrpGluGlyGlnAlaValPheLeuSerPheArgGlyGly 120
 Db 294 GCTAGGACGAGAGCTGCTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 353
 Qy 121 GlyAlaIleGluPheAlaGlnLeuMetValIleValAlaSerAlaAlaAlaArgGlyIle 140
 Db 354 GAGGCAATCGAGTTTGGACGACTGATGTTCAAGCTGCTCTAGGCTTCCAGTGGAGTT 413
 Qy 141 ProLeuGlySerValLeuThrTrpPheAspThrSerGlyLeuValIleThrValPro 160
 Db 414 CCTCTC----- 419
 Qy 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaIleTyrProIleValIleTyrGly 180
 Db 419 ----- 419
 Qy 181 ProProProProGlyTyrThrValGlnProGlyGlyGlyThrProProGlyGlyTyr 200
 Db 420 ---GAGCCCTCGGCTAT-----GATATC 440
 Qy 201 GlyAlaGlnProGlyGlyTyrGlyValAlaProProMetGlyTyrGlyValAlaProProValGly 220
 Db 441 AACCTGTATCTGAGGAGGAT---GCACT-----GTCCACCTGCT--- 479
 Qy 221 TyrGlyValProProGlyGlyTyrGlyValAlaProProGlyGlyTyrGlyValAlaProProGly 240
 Db 480 -----CCAGAGGGATGACCTGCTCCAGAGGGATGACCTGCTCGGGGG 527
 Qy 241 GlyTyrGlyAlaProProGlyGlyTyrGlyValAlaProProGlyGlyTyr-----Gly 257
 Db 528 GGCTATGCTCCCTCCGCA-----ACTCAATGATGCTTATCTTATGACCA 577
 Qy 258 AlaProProGlyGlyTyrGlyValAlaProProAlaGlyTyrGlyValAlaProProAlaGlyVal 277
 Db 578 GCTCCGATGAATGCTGACGAGCTGCTCCAGCCAGGAGGATGCTGATGCCAGAT 637
 Qy 278 GluAlaLeuProProAla 283
 Db 638 CCAGGTATGTACCCAGCG 655
 RESULT 11
 B0490729 636 bp mRNA linear EST 07-FEB-2002
 LOCUS B0490729
 DEFINITION psp2n.pk004.e6 Normalized Chicken Pituitary/Hypothalamus/Pineal
 Library (psp2n) Gallus gallus CDNA clone psp2n.pk004.e6 5' similar
 to gb|AAD10951.1 (U79458) WW domain binding protein-2 [Homo sapiens
], mRNA sequence.
 ACCESSION B0490729
 VERSION B0490729.1 GI:18611660
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 636)
Porter, T.B. and Cogburn, L.A.
ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
Library, USDA/IRAPs Animal Genome Project
Unpublished
Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickbase.udel.edu.

FEATURES
source
1. 636
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Commercial broiler chickens"
/db_xref="taxon:9031"
/clone="pgrp2n.pk004.e6"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (d12,d14,d19); post-hatch (1,3,5,7,9 weeks)"
/lab_host="B. coli BMDH108"
/clone_lib="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library (pgrp2n)"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"

BASE COUNT 143 a 174 c 173 g 146 t
ORIGIN

Alignment Scores:
Pred. No.: 2.6e-28 Length: 636
Score: 536.00 Matches: 120
Percent Similarity: 52.00% Conservative: 23
Best Local Similarity: 43.64% Mismatches: 58
Query Match: 31.36% Indels: 74
Gaps: 10

US-09-864-291-5 (1-313) x BM490729 (1-636)

QY 1 MetAlaValaInGlnSerHisrThrGluSerArgGlyAlaLeuIleProSerGlyGlu 20
Db 23 ATGGCGCTCAATAGGCGCACCTCAAG---GAAGCGGTCTCGTCCGCCAATGCCGAG 79
QY 21 SerValLeuValGlnCysGluAspValAspLeuCybPheLeuGlnYsProValGluSer 40
Db 80 AGCATTTCTAGAGATTAAGATGAGAGCTCTCTTCAGTGCATGACGAGCGAGCAGCTT 139
QY 41 TyrLeuPheAaGlyThrIleValGlyThrLeuPheLeuThrSerTyraTgValAlaPhe 60
Db 140 GAGGCGCTTCAAGGCGACCAAGGAGGAGATCTGATCTCAAGCCATACAGAGATCTTC 139
QY 61 ValThrSerHisLeuValaAspPheMetLeuSerPheMetCysProPheGlyLeuMet 80
Db 200 GTGTCAAGGGC-----AAGGATCTCATCTGCTTTATGATGACCAATTTATTTGGTG 253
QY 81 SerAspCysThrIleGluGlnProIlePheAlaProAsnTyrIleValGlyThrIleGln 100
Db 254 AAGGGGTGCTCATATGAGAGCGCTTTCTCTGCTAATATACATCAAGAGAGATTGAG 313
QY 101 AlaAlaProGlyIleGlyTyrGluGlnAlaValaPheValLeuSerPheArgGlyGly 120
Db 314 GCTGAGGCGAGAGGTGCTGAGAGGCGAGGAGAGCGTTAACTGACTTTCAACAGGGA 373
QY 121 GlyAlaIleGluPheAlaGlnLeuMetValIleAlaAspSerAlaAlaArgGlyIle 140
Db 374 GGAGCCATCGAGTTGAGACAGCTGATGTCAAGGCTGCTTAGTGCTTCACTGAGGTT 433
QY 141 ProLeuGlySerValaSerTyrTrpPheAspThrSerGlyLeuTyrIleIleThrValPro 160

Db 434 CTTCTC----- 439
QY 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGly 180
Db 439 ----- 439
QY 181 ProProProGlyIleTyrThrValGlnProGlyGluTyrGlyThrProProGlyIleTyr 200
Db 440 ---CAGCCCTTGCTAT----- 460
QY 201 GlyAlaGlnProGlyIleTyrGlyAlaProPheMetGlyTyrGlyAlaProProValGly 220
Db 461 ACACTGATACCTGAGAGGAT---GCACCT-----GTCCACCTGCT--- 499
QY 221 TyrGlyValProProGlyIleTyrGlyValProProGlyIleTyrGlyValProProGly 240
Db 500 -----CCAGAGGGGTATGACCTGCTCCAGAGGGATATGACCTGCTCGGGGG 547
QY 241 GlyTyrGlyAlaProProGlyIleTyrGlyValProProGlyIleTyrGlyAlaProPro 260
Db 548 GAGTATGCTCTCTCTCA-----CCACCTCAAAAGGT-----CCT 583
QY 261 GlyGlyTyrGlyAlaProProAla---GlyTyrGlyAlaProPro 274
Db 584 TATCTTATGACACCACTCGAATGATGATGATGATGATGATGATGATGATGATGATGAT 628

RESULT 12
BQ735296 862 bp mRNA linear Bst 16-JUL-2002
LOCUS BQ735296
DEFINITION AGENCOUNT 8096274 NIGHD_XGC_Bmb4 Xenopus laevis cDNA clone
IMAG:554398 5', mRNA sequence.
ACCESSION BQ735296
VERSION BQ735296.1 GI:21874193
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 862)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Igor David
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LML at:
www.bio.lml.gov/bdip/image/image.html
Plate: LLM12243 row: c column: 21
High quality sequence stop: 710.

FEATURES
source
1. 862
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAG:554298"
/dev_stage="embryo, stage 31-32"
/lab_host="DH10B (phage-resistant)"
/note="Organ: whole embryo; Vector: pCMV-SPORT6; Site: 1;
Nori; Site 2: Salt cloned unidirectionally. Primer: Oligo
dT. Average insert size 2.1 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC
Library)." library.

BASE COUNT 233 a 247 c 186 g 195 t 1 others
ORIGIN

Alignment Scores:

Pred. No.: 3.61e-28 Length: 862
 Score: 536.00 Matches: 121
 Percent Similarity: 54.32% Conservative: 30
 Best Local Similarity: 43.53% Mismatches: 71
 Query Match: 31.36% Indels: 56
 DB: 13 Gaps: 10

US-09-864-291-5 (1-313) x BQ735296 (1-862)

Qy 1 MetAlaValaAngInSerHisThrGluSerArgArgIyAlaLeuIleProSerGlyGlu 20
 Db 21 ATGTCACTTAATAAGAACCACTCACAGGG---GCCGGGATCATGTGTACAGACGGGAG 77
 Qy 21 SerValLeuLyGInCyseGluAspValAspLeuCysePheLeuGlnLysProValGluSer 40
 Db 78 AGTATTCACAGCAGGTAAAGATGTGAACTCTCTCAGCATATATGTCTCACAAATCA 137
 Qy 41 TyrLeuPheAenGlyThrLysLeuGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
 Db 138 GAAGCATTCAAGAGGCAAAAGAAAGATCGCTATATCTCACCGCTACAGGGGTATCTTT 197
 Qy 61 ValThrSerHisLeuValAspAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db 198 CTAAACAAGG-----AAAGACCCCATGATGATCTCTCATATGCCATTCTACCTGATG 251
 Qy 81 SerAspCystrHrIleGluGlnProIlePheAlaProAsnTyrIleGlyThrIleGln 100
 Db 252 AAAGATGCTCATAGAACAGCCAGCTCTCTGCACTACATCAAGGAACATCACTAGT 311
 Qy 101 AlaAlaProGlyGlyGlyTTPGlnGlyGlnAlaValPheLysLeuSerPheArgLysGly 120
 Db 312 GCTGAACCTGGAGATGGCTGGAGGGTAAACCTCTTTAACTTAACCTTAACAGTGA 371
 Qy 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaArgGlyLys 140
 Db 372 GGAAGCTATTTGAATTTGACAAATCATGTTCAAAATGAGCACTTGTCTTCAGACCCCT 431
 Qy 141 ProLeuGlySerValAsnTyrTrpPheAspThrSerGlyLeuTyrIleThrValPro 160
 Db 432 CCT-----GTTCCT 440
 Qy 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGly 180
 Db 441 CATGCTGCATAT----- 452
 Qy 181 ProProProProGlyTyrThrValGlnProGlyGlyTyrGlyThrProProGlyGlyTyr 200
 Db 453 -----GGATACACTCTCTGCTGCAAGAGATATGA-----CAGAGCGGATAC 494
 Qy 201 GlyAlaGlnProGlyGlyTyrGlyValaProProMetGlyTyrGlyValaProProValGly 220
 Db 495 CCTCAGGCTCCAGGAATATATACACCCCAACCAATCAAGGC-----CCTATATCA 548
 Qy 221 TyrGlyValProPro---GlyGlyTyrGlyValaProProGlyGlyTyrGlyValaProPro 239
 Db 549 TATGAGACCCCTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 605
 Qy 240 GlyGlyTyrGlyValaProProGlyGlyTyrGlyValaProProGlyGlyTyrGlyValaPro 259
 Db 606 -----TATGCTCCGCAACAGAG-----CTGGAGATGATATCACCCCT 644
 Qy 260 ProGlyGly-----TyrGlyValaProProAlaGlyTyrGlyValaProPro 274
 Db 645 CCTGAATGAATCCCATTTATCATGTGCTCTCCACACCACTTATCTGAGACCCCC 698

RESULT 13

CD301781 892 bp mRNA linear EST 27-MAY-2003
 LOCUS CD301781
 DEFINITION AGERCORT 14238680 NICHD XGC Byel Xenopus laevis cDNA clone
 IMAGE:5957790 5', mRNA sequence.
 ACCESSION CD301781
 VERSION CD301781.1 GI:31081576
 KEYWORDS EST.

SOURCE

ORGANISM

Xenopus laevis (African clawed frog)
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.

REFERENCE

1. (bases 1 to 892)
 NIH-MGC <http://mgc.ncl.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapds-remail.nih.gov

Library Procurement:
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>

Plate: L1AM14587 row: a column: 21
 High quality sequence stop: 646.

Location/Qualifiers

FEATURES

SOURCE

1..892
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:6957790"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NICHD_XGC_Byel"
 /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
 Average insert size 2.3 kb. Constructed by Life
 Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."

BASE COUNT 239 a 247 c 198 g 208 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.26e-27 Length: 892
 Score: 528.50 Matches: 125
 Percent Similarity: 52.04% Conservative: 28
 Best Local Similarity: 42.52% Mismatches: 82
 Query Match: 30.92% Indels: 59
 DB: 14 Gaps: 11

US-09-864-291-5 (1-313) x CD301781 (1-892)

Qy 1 MetAlaValaAngInSerHisThrGluSerArgArgIyAlaLeuIleProSerGlyGlu 20
 Db 41 ATGTCACTTAATAAGAACCACTCACAGGG---GCCGGGATCATGTGTACAGACGGGAG 97
 Qy 21 SerValLeuLyGInCyseGluAspValAspLeuCysePheLeuGlnLysProValGluSer 40
 Db 98 AGTATTCACAGCAGGTAAAGATGTGAACTCTCTCAGCATATATGTCTCACAAATCA 157
 Qy 41 TyrLeuPheAenGlyThrLysLeuGlyThrLeuPheLeuThrSerTyrArgValAlaPhe 60
 Db 158 GAAGCATTCAAGAGGCAAAAGAAAGATCGCTATATCTCACCGCTTACAGGATATCTTC 217
 Qy 61 ValThrSerHisLeuValAspAspProMetLeuSerPheMetMetProPheGlyLeuMet 80
 Db 218 CTAAACAAGGG-----AAAGATCCCATGATGATCTTTAATATGCACTTACATGATG 271
 Qy 81 SerAspCystrHrIleGluGlnProIlePheAlaProAsnTyrIleGlyThrIleGln 100
 Db 272 AAAGATGCTCATAGAACAGCGGTCTCTGCACTACATCAAGGAACATCACTAGT 331
 Qy 101 AlaAlaProGlyGlyGlyTTPGlnGlyGlnAlaValPheLysLeuSerPheArgLysGly 120
 Db 332 GCTGAACCTGGAGAGGGCTGGAGGGTCAAGCCTTTTAAACTTACATTTAATAGTGA 391

QY 121 GYLAALAGLuphealaglnleumetvallysalalaseralalalalarglyle 140
 DB 392 GAGCTATTGAATTGACAAATCATGTTCAAAATGACACTGTCTTCCAGAACCTT 451
 QY 141 ProleuglyserValaenlyrtrpheaapthrserserlyleuYrilelethervalpro 160
 DB 452 CCT-----GTTCTT 460
 QY 161 GYLAALAGLValCysserserGlnthrrProCyProAlaTyrrProIleValIleTyrgly 180
 DB 461 CATGCTGCATAT----- 472
 QY 181 ProProProProGlyTyrrThrrValGlnProGlyGlyTyrrThrrProProGlyGlyTyrr 200
 DB 473 -----GATACACACCTGCTCCAGAGGCTATGCA-----CATGGCGGTAC 514
 QY 201 GYLAAGlnProGlyGlyTyrrGlyValAlaProProMetGlyTyrrGlyValAlaProProValGly 220
 DB 515 CCTCAGCTCCAGAAATATATACACCCCAACCCCAACATCAGG-----CCTTATCCA 568
 QY 221 TyrrGlyValProPro---GlyGlyTyrrGlyValProProGlyGlyTyrrGlyValProPro 239
 DB 569 TATGACACCCCTGCTATGATGATGATGACCACTCCAAATCCATATGAGATACCCG--- 625
 QY 240 GlyGlyTyrrGlyValAlaProProGlyGlyTyrrGlyValProProGlyGlyTyrrGlyValAlaPro 259
 DB 626 -----TATGCTCCACAGCAGCA-----CCTGGCATATATCCACCCCT 664
 QY 260 ProGlyGly-----TyrrGlyValAlaProProAlaGlyTyrrGlyValAlaProProAlaGly 276
 DB 665 CTTGAATGAATATCCATTATACATGCTCTCCACCACTTATCTCGAAGCCCTTACAT 724
 QY 277 AaenGlnAlaLeu-----ProProAlaTyrrGlnAlaPro 287
 DB 725 GGGAGCCCTTACATGAGACCCCAACCTTCTGCTCCCA 766
 RESULT 14
 CB593075 719 bp mRNA linear EST 03-APR-2003
 LOCUS AGENCOURT 12918013 NICHD XGC Tel Xenopus laevis cDNA clone
 DEFINITION IMAGE:6899092 5', mRNA sequence.
 CB593075
 ACCESSION CB593075.1 GI:29510931
 VERSION EST.
 KEYWORDS Xenopus laevis (African clawed frog)
 ORGANISM Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 719)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Straubeberg, Ph.D.
 Email: cga@db-remail.nih.gov
 Tissue Procurement: Dr. Igor David, NICHD
 cDNA Library Preparation: CLONTECH
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNU)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNU at:
 http://image.llnl.gov
 Plate: LCM3240 row: h column: 03
 High quality sequence stop: 605.
 Location/Qualifiers
 1..719
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:6899092"
 /tissue_type="Normal testis"

/lab host="DH10B TONa"
 /clone_id="NICHD XGC Tel"
 /note="Vector: pDNR-LIB; Site 1: SfiI (directional);
 Site 2: SfiI (directional); Non-normalized full-length
 site 2: SfiI (directional); Non-normalized full-length
 enriched library 5' and 3' adaptors were used in cloning
 as follows: 5' adaptor sequence: 5'-CAGCGCATTAAGCC-3'
 and 3' adaptor sequence:
 5'-ATTCTAAGAGCCGAGGCGCCGACATG-dt(30)BN-3' (where B = A, C,
 or G and N = A, C, G, or T). Average insert size 1.2kb
 (cDNA range 0.4-3.0 kb). 15/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Corp."

BASIS COUNT 200 a 195 c 159 g 165 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6,36e-27 Length: 719
 Score: 517.00 Matches: 121
 Percent Similarity: 54.32% Conservative: 30
 Best Local Similarity: 43.53% Mismatches: 71
 Query Match: 30.25% Indels: 57
 DB: 14 Gaps: 10

US-09-864-291-5 (1-313) x CB593075 (1-719)
 QY 1 MetAlaValaenGlnserHsrHsrHsrGlyValAlaLeuIleProserGlyGly 20
 DB 39 ATGCACTTAATTAAGAACCTACAGAGG---GCCGATC-ATTGACACAAAGGAG 94
 QY 21 SerValleuylsGlnCyssGlnaPheValaPheleuCyssPheleuGlnlyPProValGlyser 40
 DB 95 AGTATCTCAGACAGTAAAGATGAGAACTCTCTCAGCAATATGCTCAAAATCA 154
 QY 41 TyrrleupheanGlnlyrlyleGlyTyrrThrrPheleuThrrSerTyrrAlaValAlaPhe 60
 DB 155 GAAGCATTCAGAGCAGAAAGAAAGAGTCCATATCTCCACCCCTACAGAGGATCTTT 214
 QY 61 ValThrrSerHsrleuValaenAPProMetleuSerPheMetPProPheGlyleuMet 80
 DB 215 CTAAAGCAAGAGG-----AAAGACCCCATATATCTTCAATATGCAATCTGATG 268
 QY 81 SerPheCyrrHsrleuGlnProIlePheAlaProaenTyrrleuGlyThrrleuGln 100
 DB 269 AAAGATGCTCCTAGACAGACAGCAGCTCTCTGCAACATCAAGGAGAACATCAGT 328
 QY 101 AlaAlaProGlyGlyTyrrGlyGlnAlaValaPheValaPheleuSerPheAlaGlyGly 120
 DB 329 GCTGAACCTGAGATGCTGGAGGGTAAACCTCTTTAAACCTTCAATTAAACAGTGA 388
 QY 121 GYLAALAGLuphealaglnleumetvallysalalaseralalalalarglyle 140
 DB 389 GAGGCTATTGAATTGACAAATCATGTTCAAAATGACACTTCTTCCAGAACCTT 448
 QY 141 ProleuglyserValaenlyrtrpheaapthrserserlyleuYrilelethervalpro 160
 DB 449 CCT-----GTTCTT 457
 QY 161 GYLAALAGLValCysserserGlnthrrProCyProAlaTyrrProIleValIleTyrgly 180
 DB 458 CATGCTGCATAT----- 469
 QY 181 ProProProProGlyTyrrThrrValGlnProGlyGlyTyrrThrrProProGlyGlyTyrr 200
 DB 470 -----GATACACTCTGCTCCAGAGATATGCA-----CATGGCGGTAC 511
 QY 201 GYLAAGlnProGlyGlyTyrrGlyValAlaProProMetGlyTyrrGlyValAlaProProValGly 220
 DB 512 CCTCAGCTCCAGAAATATATACACCCCAACCCCAACATCAGG-----CCTTATCCA 565
 QY 221 TyrrGlyValProPro---GlyGlyTyrrGlyValProProGlyGlyTyrrGlyValProPro 239
 DB 566 TATGACACCCCTGCTATGATGATGATGACCACTCCAAATGCTATGAGATACCCG--- 622

QY 240 GlycylTyrGlyAlaProGlyGlyTyrGlyValProGlyGlyTyrGlyAlaPro 259
 DB 623 -----TATGCTCCGCCACACAGG-----CCTGGCATGATCCACCCCT 661
 QY 260 ProGlyGly-----TyrGlyAlaProProAlaGlyTyrGlyAlaProPro 274
 DB 662 CCTGAATGAATCCATTACATGAGCTCTCCACACCTTATCTGAGCCCG 715
 RESULT 15
 AL898723
 LOCUS AL898723 XGC-egg Silurana tropicalis cDNA clone TSG9036a07 5', mRNA
 DEFINITION AL898723 XGC-egg Silurana tropicalis cDNA clone TSG9036a07 5', mRNA
 ACCESSION AL898723
 VERSION AL898723
 KEYWORDS AL898723.1 GI:22951115
 SOURCE EST
 ORGANISM Silurana tropicalis (western clawed frog)
 Silurana tropicalis
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Silurana.
 1 (bases 1 to 655)
 Taylor, R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.
 Sanger Xenopus tropicalis EST project 2002
 JOURNAL Unpublished
 COMMENT Contact: Taylor R
 Sanger Centre
 Hinxton, Cambridgeshire, CB10 1SA, UK
 Email: trop@sanger.ac.uk
 Sanger Xenopus tropicalis EST project 2001
 TROPICALIS SEQUENCE ID: TSG9036a07.plkSP6
 Sequencing primer: SP6
 This sequence is from a Xenopus Gene Collection (XGC) library
 constructed by Aaron M. Zorn.
 FEATURES
 SOURCE
 1..655
 Location/Qualifiers
 /organism="Silurana tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="TSG9036a07"
 /dev_stage="egg"
 /lab_host="Escherichia coli XL1-blue"
 /clone_1lb="XGC-egg"
 /note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
 was oligo dt primed from 5ug of poly A+ RNA from egg.
 EcoRI-NotI cut cDNA was then ligated into pCS107 with
 EcoRI at the 5' end and NotI at the 3' end"
 BASE COUNT 172 a 172 c 155 g 156 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.1e-26 Length: 655
 Score: 513.00 Matches: 112
 Percent Similarity: 55.82% Conservative: 27
 Best Local Similarity: 44.98% Mismatches: 62
 Query Match: 30.02% Indels: 48
 DB: 9 Gaps: 8
 US-09-864-291-5 (1-313) x AL898723 (1-655)
 QY 1 MetAlaValAlaSerIleThrGluSerArgGlyAlaIleuIleProSerGlyIu 20
 DB 39 ATGTCCCTAAATTAAGACCACTCTGAGGG---GCTGGATCATCTCAGCAACGGGAG 95
 QY 21 SerValLeuLysGlnCysGluAlaPValAlaPLeuCyPheLeuGlnLysProValGluSer 40
 DB 96 AGTGTCTCAGCAGAGTAAAGATGTGGAATCTCTCTCAGCAGATATGTCTCAAAAATCA 155
 QY 41 TyrLeuPheAsnGlyThrIleLysGlyThrIlePheLeuThrSerTyrArgValAlaPhe 60
 DB 156 GAGGATTCAAAGGACCAAAAAGATCATCTTTCTCAACCCCTACAGGGGTGATCTTC 215

QY 61 ValThrSerHisLeuValaAsnProMetLeuSerPheMetMetProPheGlyLeuMet 80
 DB 216 CTAAAGCAAGGGG-----AAGAGCCCATATGATCTTTATATGACATTTATCTGATG 269
 QY 81 SerAspCysThrIleGluGlnProIlePheAlaProAsnTyrIleLysGlyThrIleGln 100
 DB 270 AAAGATGCTTCATAGAACACCCAGTGTCTCTCCACATACATCAAGGATCAATCGGT 329
 QY 101 AlaAlaProGlyGlyGlyTyrGluGlyGlyAlaValAlaPheLysLeuSerPheArgLysGly 120
 DB 330 GCTAGGCTGAAGGTGCTGGGTGGGTCAAGCCCTTTCAAACTTACATTTAATAGTGA 389
 QY 121 GlyAlaIleGluPheAlaGlnLeuMetValLysAlaAlaSerAlaAlaAlaGlyTyrIle 140
 DB 390 GGTGCTATTGAATTGACAAAGTCATGTTCAAAATGCAACTGTGCTTCCAGA----- 443
 QY 141 ProLeuGlySerValaAsnTyrTrpPheAspThrSerGlyLeuTyrIleIleThrValPro 160
 DB 443 ----- 443
 QY 161 GlyAlaAlaValCysSerSerGlnThrProCysProAlaTyrProIleValIleTyrGly 180
 DB 444 -----GCCCTCTGTCCTCCTCATGCCGTATATGGA 473
 QY 181 -----ProProProProGlyTyrThrValGlnProGlyGlyTyrThrProGly 198
 DB 474 TACATTCCTGCACAGCGGGTACCTCCAGCTCCAGGAATATATTCAGCCGCCCA 533
 QY 199 GlyTyrGlyAlaGlnProGlyGlyTyrGlyAlaProMet---GlyTyrGlyAlaPro 217
 DB 534 CCATCAGGACCTTATCCA-----TACGAGCCCCCTGCTGAATGATATGAGCCACCT 587
 QY 218 ProValGlyTyrGlyValProProGlyGlyTyrGlyValProProGlyGlyTyrGlyVal 237
 DB 588 CCAATTCAAATGGATATACCC-----TATGCTCAGCACACAGGA----- 626
 QY 238 ProProGlyGlyTyrGlyAlaProPro 246
 DB 627 ---CCTGGTATGATCCACCCCTCTCT 650

Search completed: December 16, 2003, 20:45:08
 Job time : 2048.43 secs